

**IJCSIS Vol. 8 No. 9, December 2010**  
**ISSN 1947-5500**

# **International Journal of Computer Science & Information Security**

**© IJCSIS PUBLICATION 2010**

## Editorial

### Message from Managing Editor

*The International Journal of Computer Science and Information Security (IJCSIS) is a well-established publication venue on novel research in computer science and information security. The year 2010 has been very eventful and encouraging for all IJCSIS authors/researchers and IJCSIS technical committee, as we see more and more interest in IJCSIS research publications. IJCSIS is now empowered by over thousands of academics, researchers, authors/reviewers/students and research organizations. Reaching this milestone would not have been possible without the support, feedback, and continuous engagement of our authors and reviewers.*

*Field coverage includes: security infrastructures, network security: Internet security, content protection, cryptography, steganography and formal methods in information security; multimedia systems, software, information systems, intelligent systems, web services, data mining, wireless communication, networking and technologies, innovation technology and management. ( See monthly Call for Papers)*

*We are grateful to our reviewers for providing valuable comments. IJCSIS December 2010 issue (Vol. 8, No. 9) has paper acceptance rate of nearly 35%.*

*We wish everyone a successful scientific research year on 2011.*

Available at <http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 8, No. 9, December 2010 Edition

ISSN 1947-5500 © IJCSIS, USA.

*Abstracts Indexed by (among others):*



## IJCSIS EDITORIAL BOARD

**Dr. Gregorio Martinez Perez**

Associate Professor - Professor Titular de Universidad, University of Murcia (UMU), Spain

**Dr. M. Emre Celebi,**

Assistant Professor, Department of Computer Science, Louisiana State University in Shreveport, USA

**Dr. Yong Li**

School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China

**Prof. Hamid Reza Naji**

Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran

**Dr. Sanjay Jasola**

Professor and Dean, School of Information and Communication Technology, Gautam Buddha University

**Dr Riktesh Srivastava**

Assistant Professor, Information Systems, Skyline University College, University City of Sharjah, Sharjah, PO 1797, UAE

**Dr. Siddhivinayak Kulkarni**

University of Ballarat, Ballarat, Victoria, Australia

**Professor (Dr) Mokhtar Beldjehem**

Sainte-Anne University, Halifax, NS, Canada

**Dr. Alex Pappachen James, (Research Fellow)**

Queensland Micro-nanotechnology center, Griffith University, Australia

Dr. T.C. Manjunath,

ATRIA Institute of Tech, India.

## TABLE OF CONTENTS

### 1. Paper 23111015: Chaotic Hash Function Based on MD5 and SHA-1 Hash Algorithms (pp. 1-3)

*Mouad HAMRI & Jilali Mikram,*

*Applied Mathematics and computer science department, Science University of Rabat, 4 Avenue Ibn Battouta Rabat Morocco*

*Fouad Zinoun, Economical sciences and management department, University of Meknes*

**Abstract** — In this work an implementation of a 128 bits chaotic hash function is presented. We have used the logistic map as the chaotic dynamical system of our algorithm and for the initial condition generation of the chaotic map, we've used the two hash functions MD5 and SHA-1. The main idea is to use the MD5 and SHA-1 digits of the a message to construct the initial condition of the logistic map and using this map we will generate the digit of this message. The algorithm was implemented using JAVA programming language and its performance was studied by comparing it to MD5 hash function.

**Keywords:** *Cryptography, Hash functions, MD5, SHA, Dynamical systems, Chaos theory.*

### 2. Paper 23111018: Concept Measure Supported Aesthetics Ontology Construction for Videos (pp. 4-7)

*Dr Sunitha Abburu, Professor & Director, Department of Computer Applications, Adhiyamaan College of Engineering, Hosur, pin-635109, Tamilnadu, India*

**Abstract** — Entertainment plays a vital role in human life. Multimedia conquers top position in the entertainment world. Video stands first among the multimedia entertainment. The rapid growth of the videos has resulted in the emergence of numerous multimedia repositories that require efficient and effective video storage, semantic annotation, indexing and retrieval systems. The introduction of ontologies in multimedia retrieval system can improve the precision and recall rate effectively. The performance of the annotation and the retrieval system increases with the support of the domain ontology. Most of the video annotation, indexing and the retrieval systems focus on the semantic concepts like objects, people, location, events, actions etc. But most of the multimedia systems are filled with human and their emotions. Any multimedia system like cinema, news videos, sports videos, and any domestic functional videos tries to capture the emotions of the human involved in the occasion. A video retrieval system will be complete if the system identifies, captures and represents the emotions of the humans. This paper focus on identification and representation of human emotions and the intensity of the emotions are represented using the fuzzy logic. The concept of Navarasra has been brought in to video retrieval system to classify the human emotions. The proposed approach is generic and flexible. It is designed and constructed for all videos where human beings and their emotions are been captured. A practical implementation is done using Protégé as an Ontology developing tool.

**Keywords-component;** *Video Semantics, Concept measures, Ontology, Retrieval, Human Emotions.*

### 3. Paper 26111027: Reconfigurable Hardware Intelligent Memory Controller for H.264/AVC Encoders (pp. 8-16)

*Kamel Messaoudi, LERICA Laboratory Badj Mokhtar University Annaba, Algeria*

*El-Bay Bourennane, LE2I Laboratory, Burgundy University, Dijon, France*

*Salah Toumi, LERICA Laboratory, Badj Mokhtar University, Annaba, Algeria*

**Abstract** — In this paper, we propose an intelligent memory controller for H.264/AVC CODECs with memory management strategies based on a preloading mechanism in order to reduce the number of accesses to external memory. The controller is used both in simulation and implementation platform for the H.264 encoder. This platform uses an external DDR2 memory to record an image sequence, and an



intelligent component with local memories to read the images periodically according to the needs of the H.264 processing modules. Several on-chip memories are used to avoid accessing to off-chip memory and consequently decrease energy consumption. The proposed memory controller can be adapted to the various profiles defined in the standard. The intelligent controller is implemented in VHDL and verified to run at 114 MHz using a Xilinx virtex5-ML501 platform. The proposed architecture occupies 10% of the FPGA's resources and ensures data for the processing modules in H.264 encoder.

*Keywords- Memory management; intelligent controller; FPGA; H.264/AVC codec; real-time processing.*

#### **4. Paper 26111030: Fragmentation Investigation And Evaluation In Distributed DBMS Using JDBC and OGSA-DAI (pp. 17-24)**

*Ahmed Almadi, Ahmed Manasrah, Omer Abouabdalla, Homam El-Taj*

*National Advanced IPv6 Center of Excellence (NAv6) Universiti Sains Malaysia Penang, Malaysia*

*Abstract* — This research investigates and evaluate the impact of the fragmentation on different database retrieval modes based on derived horizontal fragmentation by generating and distributing the query to the servers (distributed search) or send the query to the direct server (direct search). Moreover, it provides recommendation on suitable query execution strategies based on a proposed fitness fragmentation formula. Furthermore, examine the suitable technology such as OGSA-DAI and JDBC in grid database to examine the time overhead in distributed systems and grid environments in different cases like size or number of servers. The results show that the fragmentation's time performance impact is clearly effective and positively applied while increasing the database size or the number of servers. On the other hand, the OGSA-DAI kept on showing slower execution time on all conducted scenarios, and the differences between the execution time exceeds up to 70% while increasing the size of data or number of servers. In addition, this thesis has tested the impact of fragmentation search against the distributed search where the first one submit the query to direct server(s) (direct search), and the second one distribute the query to the servers (distributed search). The result shows that the speed effectiveness of direct search technique in JDBC case is around 70% faster than the distributed search and around 50% faster in OGSA-DAI case.

*Keywords-component; JDBC; OGSA-DAI; Fragmentation; Distributed DBMS*

#### **5. Paper 29111036: Impact of Guard Interval in Proposed MIMO-OFDM System for wireless communication (pp. 25-30)**

*M. P. Chitra, Research Scholar, Sathyabama University, Chennai, India.*

*Dr. S. K. Srivatsa, Senior Professor, St. Joseph College of Engineering, Chennai, India.*

*Abstract* - Alamouti's space-time coding scheme for Multi-Input Multi-Output (MIMO) system has drawn much attention in 4G wireless technologies. Orthogonal frequency division multiplexing (OFDM) is a popular method for high data rate wireless transmission. OFDM may be combined with antenna arrays at the transmitter and receiver to increase the diversity gain and enhance the system capacity on time variant and frequency selective channels, resulting in Multi-Input Multi-Output (MIMO) configuration. This paper explores various physical layer research challenges in MIMO-OFDM system design including channel modeling, space time block code techniques, channel estimation and signal processing algorithms used for performing time and frequency synchronization in MIMO-OFDM system. The proposed system is simulated in matlab and analyzed in terms of BER with signals to noise ratio (SNR). The difference of BER for coded and uncoded MIMO system and also the impact of guard interval are simulated using different wireless channel.

*Keywords - Multi-Input Multi-Output (MIMO); orthogonal frequency division multiplexing (OFDM); Bit error rate (BER); signals to noise ratio (SNR); Single input single output (SISO); space time block code (STBC)*

## **6. Paper 29111037: Implementation of New Routing Protocol for Node Security in a Mobile Ad Hoc Network (pp. 31-36)**

*Virendra Singh Kushwah, Department of Computer Science, Hindustan Institute of Management and Computer Studies, Farah, Mathura, INDIA*

*Gaurav Sharma, Department of Computer Science, GLA, University Mathura, INDIA*

**Abstract--** A routing protocol plays important role to handle entire network for communication and determines the paths of packets. A node is a part of the defined network for transferring information in form of packets. If all packets transferred from source to destination successfully, it has been assumed that the routing protocol is good. But, an attacker turns this dealing as a speed breaker and turning point of a highway. So, prevention from attacks and secure packets, a new routing protocol is being introduced in this paper. The proposed routing protocol is called by SNAODV (Secure Node AODV). This paper is also tried to maximize throughput as compared with AODV and SAODV.

**Keywords—**AODV; routing; packets; network

## **7. Paper 30111038: Runtime Monitoring and Controlling Of Information Flow (pp. 37-45)**

*Mohamed Sarrah, Software Technology Research Laboratory, De Montfort University, Leicester, LE1 9BH, UK*

*Helge Janicke, Software Technology Research Laboratory, De Montfort University, Leicester, LE1 9BH, UK*

**Abstract—** Computer systems are monitored to check performance or verified to check the correctness of the software systems with respect to security properties such as integrity, availability and confidentiality. The more sensitive the information, such as credit card information, military information or personal medical information, being processed by the software, the more necessary and important it is to monitor and control the flow of sensitive information. Monitoring and controlling an untrusted program behavior to ensure information confidentiality during runtime in an environment where confidential information is present is often difficult and unnerving task for users. The issue is how users can monitor and control the flow of the confidential information at the runtime of untrusted programs. In this paper we present a novel runtime verification approach for monitoring and controlling information flow that supports user interaction with a running program.

**Keywords-** Information flow control; Runtime monitoring; Confidentiality; Usability.

## **8. Paper 30111047: Design of Novel Usability Driven Software Process Model (pp. 46-53)**

*Hina Gull, Department of Computer Engineering, College of Electrical & Mechanical Engineering NUST, Rawalpindi Pakistan*

*Farooque Azam, Department of Computer Engineering, College of Electrical & Mechanical Engineering NUST, Rawalpindi Pakistan*

*Sardar Zafar Iqbal, Department of Computer Sciences, Iqra University, Islamabad Pakistan*

**Abstract -** In this paper we have proposed a novel software process model for web based applications. This model is based on the empirical study carried out by us and also by the literature review of software development models. Model consists of three phases: Requirement Engineering, Design and Implementation. Model contains certain sub activities in each phase describing the flow and steps which should be followed to develop a web application. The main emphasis of the model is on usability aspect, keeping in view the criticality of the user interface for a good web application. Flexible and easy change manageable nature of the model makes it different and worth using as compared to other software development approaches.

### **9. Paper 30111050: Power-Based Key Hopping (PBKH) and Associated Hardware Implementation (pp. 54-60)**

*Rabie A. Mahmoud, Department of Mathematics, Faculty of Science, Cairo University, Cairo, Egypt.  
Magdy Saeb, Computer Engineering Department, Arab Academy for Science, Tech. & Maritime Transport,  
Alexandria, Egypt.*

**Abstract:** Power-Based Key Hopping (PBKH) is a process of key hopping that is used to interchange the user key of a cipher. Power based key hopping is founded on the idea of dynamic frequency power-based hopping to change the user key. This is achieved through computing the power of a previous cipher packet and comparing it with a standard value. In this work, we discuss various key hopping methods and suggest a procedure of power based key hopping. Moreover, we provide a Field Programmable Gate Array (FPGA) hardware implementation of the proposed key hopping technique.

**Keywords:** Power Based Key Hopping; Security; Hardware; FPGA

### **10. Paper 30111058: The Innovative Application of Multiple Correlation Plane (pp. 61-69)**

*Julaluk Watthananon, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand  
Sageemas Na Wichian, College of Industrial Technology, King Mongkut's University of Technology North Bangkok, Thailand  
Anirach Mingkhwan, Faculty of Industrial and Technology Management, King Mongkut's University of Technology North Bangkok, Thailand*

**Abstract—**Presentation data with column graph and line graph is a well-known technique used in data explanation to compare and show direction that users can easily understand. However, the techniques has limitations on the data describing complex with multiple relations, that is, if the data contains diverse relationships and many variables, the efficiency of the presentation will decrease. In this paper, the mathematical method for multi relations based on Radar graph is proposed. The position of information approaches on the correlation plane referred to the distribution of content and the deep specific content. However, the proposed method analyzes the multi variants data by plotting in the correlation plane, and compared with the base line system. The result shows that the performance is higher than other methods in term of accuracy, time and features.

**Keywords-**Correlation plane; correlation boundary; correlation plot; Star plot; Radar graph

### **11. Paper 30111059: Structural Analysis of Bangla Sentences of Different Tenses for Automatic Bangla Machine Translator (pp. 70-75)**

*Md. Musfique Anwar, Nasrin Sultana Shume and Md. Al-Amin Bhuiyan  
Dept. of Computer Science & Engineering, Jahangirnagar University, Dakha, Bangladesh*

**Abstract -** This paper addresses about structural mappings of Bangla sentences of different tenses for machine translation (MT). Machine translation requires analysis, transfer and generation steps to produce target language output from a source language input. Structural representation of Bangla sentences encodes the information of Bangla sentences and a transfer module has been designed that can generate English sentences using Context Free Grammar (CFG). The MT system generates parse tree according to the parse rules and a lexicon provides the properties of the word and its meaning in the target language. The MT system can be extendable to paragraph translation.

**Keywords:** Machine Translation, Structural representation, Context Free Grammar, Parse tree, Lexicon etc.

## **12. Paper 30111064: Image Retrieval using Shape Texture Patterns generated from Walsh-Hadamard Transform and Gradient Image Bitmap (pp. 76-82)**

*Dr. H.B.Kekre, Sudeep D. Thepade, Varun K. Banura*

*Computer Engineering Department, MPSTME, SVKM's NMIMS (Deemed-to-be University), Mumbai, India*

**Abstract** — The theme of the work presented here is gradient mask texture based image retrieval techniques using image bitmaps and texture patterns generated using Walsh-Hadamard transform. The shape of the image is extracted by using three different gradient operators (Prewitt, Robert and Sobel) with slope magnitude method followed by generation of bitmap of the shape feature extracted. This bitmap is then compared with the different texture patterns namely '4-pattern', '16-pattern' and '64-pattern' generated using Walsh-Hadamard transform matrix to produce the feature vector as the matching number of ones and minus ones per texture pattern. The proposed content based image retrieval (CBIR) techniques are tested on a generic image database having 1000 images spread across 11 categories. For each proposed CBIR technique 55 queries (randomly selected 5 per image category) are fired on the image database. To compare the performance of image retrieval techniques average precision and recall of all the queries per image retrieval technique are computed. The results have shown improved performance (higher precision and recall values of crossover points) with the proposed methods compared to the mask-shape based image retrieval techniques. Further the performance of proposed image retrieval methods is enhanced using even image part. In the discussed image retrieval methods, the combination of original and even image part for 4-pattern texture with shape masks generated using Robert gradient operator gives the highest crossover point of precision and recall indicating better performance.

**Keywords-** CBIR, Gradient operators, Walsh-Hadamard transform, Texture, Pattern, Bitmap.

## **13. Paper 30111075: An Evaluation of Software Requirement Prioritization Techniques (pp. 83-94)**

*Mohammad Shabbir Hasan, Abdullah Al Mahmood, Md. Jahangir Alam, Sk. Md. Nahid Hasan, Farin Rahman*

*Panacea Research Lab, Dhaka, Bangladesh and Dhaka International University, Dhaka, Bangladesh.*

**Abstract** — Requirements prioritization plays an important role in the requirement engineering process, particularly, with respect to critical tasks like requirements negotiation and software release planning. Selecting the right set of requirements for a product release largely depends on how successfully the requirement candidates are prioritized. There are different requirement prioritization techniques available which are some more elaborated than others. This paper takes a closer look at nine different techniques of requirement prioritization namely Analytical Hierarchy Process (AHP), Hierarchy AHP, Minimal Spanning Tree, Bubble Sort, Binary Search Tree (BST), Priority Group, Planning Game (PG), 100 points method and Planning Game combined with AHP (PGcAHP) and then put them into a controlled experiment, in order to find out the best one. The evaluation was done on the basis of some criteria like: ease of use, certainty, accuracy of result, method's ability to scale up to many more requirements, required number of comparisons, and required time to make decision. Analysis of the data from the experiment indicates that the analytic hierarchy process to be a promising candidate, although it may be problematic to scale up. However, the result clearly indicates that the Planning Game (PG) yields accurate result, is able to scale up, requires least amount of time, the easiest method to use and so on. For these reasons, finding of the experiment is, the Planning Game (PG) method is supposed to be the best method for prioritizing requirements.

**Keywords-** Requirement Engineering, Requirement Prioritization, Requirement Negotiation, Software Product Management, Software Release Planning.



**14. Paper 30111076: Routing Fairness in Mobile Ad-Hoc Networks: Analysis and Enhancement (pp. 95-100)**

*M. Shiraz Ahmed, Lecturer, Department Of Computer Science & Information Systems, Najran University, Najran, Saudi Arabia,*  
*Haji Moinuddin, Lecturer, Department Of Computer Science & Information Systems, Najran University, Najran, Saudi Arabia,*  
*Fazeelatunnisa, Lecturer, Department Of Computer Science & Information Systems, Najran University, Najran, Saudi Arabia*

**Abstract:** With the rapid advances in wireless and semiconductor technologies mobile connectivity became cheap and ubiquitous. One of the major challenges facing Mobile Ad-Hoc Networks (also known as MANETs) is the absence of a proper routing protocol that provides good fairness and scalability, low overhead, low end-to-end delays, seamless connectivity and good quality of service. This paper studies the fairness of routing protocols for MANETS. In this paper we propose routing segments methods to solve the problem of lack of fairness in routing.

**Keywords:** MANETS, Fairness, Segments, Scalability

**15. Paper 30111054: Nano-particle Characterization Using a Fast Hybrid Clustering Technique for TEM Images (pp. 101-110)**

*M.A. Abdou, Researcher, IRI, Mubarak City for Scientific Research and Technology Applications, Alexandria, Egypt.*  
*Bayumy B.A. Youssef, Researcher, IRI, Mubarak City for Scientific Research and Technology Applications, Alexandria, Egypt.*  
*W.M. Sheta, Researcher, IRI, Mubarak City for Scientific Research and Technology Applications, Alexandria, Egypt*

**Abstract -** This Paper introduces a new fast Transmission Electron Microscopy (TEM) images clustering technique. Since analysis of particle sizes and shapes from two-dimensional TEM images is affected by variations in image contrast between adjacent particles, automatic methods requires more efforts. The proposed hybrid method consists of two main steps: automatic segmentation and nano-particles counting. The segmentation procedure begins with an automatic threshold generator and moves towards a high efficient multiple- regions segmentation technique. Results are observed, compared with existing methods and manual counting.

**Keywords:** TEM, Image segmentation, Threshold generator, Nano-particle counting

**16. Paper 30111042: Gaussian Process Model for Uncertain Data Classification (pp. 111-115)**

*G.V. Suresh, Assoc. Professor, CSE Department, Universal College of Engineering, Guntur, India*  
*Shabbeer Shaik, Assoc. Professor, MCA Department, Tirmula College of Engineering , Guntur, India*  
*E.V.Reddy, Assoc. Professor, CSE Department, Universal College of Engineering, Guntur, India*  
*Usman Ali Shaik, Assoc. Professor, CSE Department, Universal College of Engineering, Guntur, India*

**Abstract—** Data uncertainty is common in real-world applications due to various causes, including imprecise measurement, network latency, out-dated sources and sampling errors. These kinds of uncertainty have to be handled cautiously, or else the mining results could be unreliable or even wrong. We propose that when data mining is performed on uncertain data, data uncertainty has to be considered in order to obtain high quality data mining results. In this paper we study how uncertainty can be incorporated in data mining by using data clustering as a motivating example. We also present a Gaussian process model that can be able to handle data uncertainty in data mining.

**Keywords-** Gaussian process, uncertain data, Gaussian distribution, Data Mining

### **17. Paper 30111078: Pair Wise Sorting: A New Way of Sorting (pp. 116-120)**

*Md. Jahangir Alam, Dhaka International University, Dhaka, Bangladesh*

*Muhammad Monsur Uddin, Institute of Science, Trade and Technology, Dhaka, Bangladesh*

*Mohammad Shabbir Hasan, Abdullah Al Mahmood, Panacea Research Lab, Dhaka, Bangladesh*

**Abstract** — This paper presents a technique for sorting numerical data in an efficient way. The numbers of comparisons i.e. the running time of this technique is dependent on distribution or diversity of the value of data items as like as other efficient algorithms. When the total number of data is even, this method groups that data into a collection of pairs and therefore establishes the sorting constraints on each of the pairs. The control is traversed through the list of elements by changing the position of each pair which is the major principle of this technique. On the other hand, when the total number of elements is odd, this method sorts all elements except the last one in the same way as mentioned earlier and the last element is sorted using the general Insertion Sort. This algorithm is therefore a hybrid sorting method that sorts elementary numeric data in a faster and efficient manner.

**Keywords-** *Sorting, Pair Wise Sorting, Sorting Techniques.*

### **18. Paper 30111086: Improving Client-Server Response Time Using IPv6 Header Suppression Over MPLS (pp. 121-126)**

*Imad J. Mohamad & Yung-Wey Chong, National Advanced IPv6 Center, Universiti Sains Malaysia, 11800 Penang, Malaysia*

*Tat-Chee Wan & Putra Sumari, School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia*

**Abstract** — Optimizing the response time for Client-Server IPv6 traffic over label switched path (LSP) is the main contribution for this paper. It is achieved using header suppression for real time IPv6 traffic across MPLS label switched path (LSP). Robust Header Compression (RoHC) and Payload Header Suppression (PHS) are two options defined in IEEE 802.16 for mobile WiMAX performance work using link-by-link approach. This paper adapts PHS for MPLS performance and extends it to work over LSP using end-to-end approach. The implementation for IPv6 header suppression using NS2 shows improvement in response time for client-server traffic by 1.7s. Additional improvement in QoS parameters for UDP and TCP traffic is investigated.

**Keywords-component;** *Client-Server Traffic, LSP, IPv6, Header Suppression;*

### **19. Paper 31101070: Hybrid Compression of Color Images with Larger Trivial Background by Histogram Segmentation (pp. 127-131)**

*M. Mohamed Sathik, Department of Computer Science, Sadakathullah Appa College, Tirunelveli, India*

*K. Senthamarai Kannan and Y. Jacob Vetha Raj, Department of Statistics, Manonmanium Sundaranar University, Tirunelveli, India*

**Abstract** - A hybrid image compression method is proposed by which the background of the image is compressed using lossy compression and the rest of the image is compressed using lossless compression. In Hybrid Compression of Color Images with Larger Trivial Background by Histogram Segmentation(HCCILTBHS) , input color image is subjected to binary segmentation using histogram to detect the background. The color image is compressed by standard lossy compression method. The difference between the lossy image and the original image is computed and is called as residue. The residue at the background area is dropped and rest of the area is compressed by standard lossless compression method. This method gives lower bit rate than the lossless compression methods and is well suited to any color image with larger trivial background.

**Key Words-** *Segmentation, Erosion, Dilation, Image Compression.*

**20. Paper 30101074: Realization and Study of High Performance Voltage Mode Oscillator based on CCCCTA: A Building Block for Analog Signal Processing (pp. 132-138)**

*Deependra Pandey, Asst. Professor, Dept. of ECE, Amity University, Lucknow  
Prof.(Dr.) L.K. Singh, Director, IET, Dr. R. M. L. Avadh University, Faizabad*

**Abstract-** At present there is a growing interest in designing current mode circuits. This attributed to their large signal bandwidth, great linearity, wider dynamic range, simple circuitry and low power consumption. The paper presents a basic current-mode building block for analog signal processing, namely current controlled current conveyor transconductance amplifier (CCCCTA). Its parasitic resistance at current input port can be controlled by an input bias current. It is very suitable to use in a current-mode signal processing, which is continually more popular than a voltage one. The proposed element is realized in a CMOS technology and is examined the performances through PSPICE simulations. The CCCCTA performs tuning over a wide current range. In addition, some circuits for example as a current-mode universal biquad filter and a grounded inductance occupy only single CCCCTA.

**Keywords:** *Current Conveyors, CCCCTA, Current-mode circuits, Voltage Mode Oscillator*

**21. Paper 11101002: Hybrid Technique for Self Tuning PI Controller Parameters in HVDC Systems (pp. 139-150)**

*A.Srujana, Research Scholar, JNT University, Hyderabad  
Dr. S.V.Jayaram Kumar, Professor, Jawaharlal Nehru Technological University, Hyderabad*

**Abstract** — Nowadays, due to certain advantages, the HVDC systems are commonly used in long distance transmissions. The major drawback associated with HVDC system is that it takes a longer duration to return to its steady state value after the occurrence of a fault. In a HVDC system, when a fault occurs, the current and voltage will deviate from their normal range and PI controllers are used to maintain its current and voltage at the normal steady state value. Controller parameter tuning plays a significant role in maintaining the steady state current and voltage of a HVDC system. Here, we propose a hybrid technique to self tune the PI controller parameters. The proposed hybrid technique utilizes fuzzy logic and neural network to self tune the controller parameters. The fuzzy rules are generated using different combinations of current error, rate and combined gain. To train the neural network, different combinations of fuzzy gain, proportional gain and integral gain are used. The neural network is trained using a back propagation algorithm. By experimentation it is shown that the system that uses this method takes a very short time to return to its normal steady state. The implementation results show that the performance of the proposed hybrid technique is superior to that of both the self tuning techniques.

**Keywords-** *fuzzy logic; HVDC; neural network; fuzzy rules; proportional and integral gain.*

**22. Paper 26111032: Use of Computerized Web-Based Information System For Determining Losses in 15-6.6 KV Feeders in Traditional Electrical Network Management: Case Study Goma Distribution Electrical Network (pp. 151-157)**

*Ezekiel U. Okike, Department of Computer Science, University of Ibadan, Ibadan, Nigeria  
Bakunzi G. Joseph, School of Computer Studies, Kampala International University, Kampala, Uganda*

**Abstract**— Electrical energy plays very vital role in modern global economy. The aim of this study is to develop a framework for a Web-Based Information System (WIS) tool for computing losses from 15 – 6.6 KV Feeders in Traditional Electrical Network Management (TENM). The study was conducted in Goma District in the Democratic Republic of Congo. Data were collected from 26 key staff of Goma Distribution Electrical Network who responded to the questionnaires and from metered reading documents used in the study. The study implemented a Computerized Web-Based Information System (CWIS) to compute different losses in Goma electrical distribution network. The CWIS computed technical losses in five 15-6.6KV feeders of Goma electrical distribution network. The study revealed that among the five feeders, feeder 1 (Sud feeder) consumes 1,469,172.6 KWH representing 66.3% of the total annual energy loss while



others presented lower annual losses. This is an indication that Feeder 1 is overloaded and needed to be resized or on the alternative, the installation of another overhead cable that will take the half of the load in charge.

*Keywords- Electrical energy; energy distribution; feeder loss; computerized information system*

### **23. Paper 27101035: A Design and Execution of Activity Based Applications In Distributed Environments Using Hermes Software Architecture (pp. 158-166)**

*B. Muthukumar, P. Banumathi & T.K.P. Rajagopal  
Kathir College of Engineering, Coimbatore, Tamilnadu, INDIA  
R.Vidhya, PG Student, Pondicherry University, Pondicherry*

*Abstract* - Hermes is an agent-based middleware structured as component-based and 3-layered software architecture. Hermes provides an integrated, flexible programming environment for design and execution of activity-based applications in distributed environments. By using workflow technology, it supports even a non expert user programmer in the model driven design and implementation of a domain specific application. In this paper, after a description of Hermes software architecture, we provide a simple demo in biological domain and we show some real case studies in which Hermes has been validated.

*Keywords:* Hermes Software Architecture, O2I Project, Agents, Run –Time Layers etc.

### **24. Paper 27101036: Tracking The Face Movement Of Perceptual User Interface Using CAMSHIFT Algorithm (pp. 167-175)**

*B. Muthukumar, Assistant Professor, Information Technology, Kathir College of Engineering, Coimbatore  
Dr. S. Ravi, Professor and Head, ECE Department, Dr.M.G.R University, Chennai*

*Abstract:* This paper deals with a perceptual user interface and computer vision color tracking algorithm is developed and applied towards tracking human faces. Computer vision algorithms that are intended to form part of a perceptual user interface must be fast and efficient. They must be able to track in real time yet not absorb a major share of computational resources; other tasks must be able to run while the visual interface is being used. The new algorithm developed here is based on a robust nonparametric technique for climbing density gradients to find the mode (peak) of probability distributions called the mean shift algorithm. In our case, we want to find the mode of a color distribution within a video scene. Therefore, the mean shift algorithm is modified to deal with dynamically changing color probability distributions derived from video frame sequences. The modified algorithm is called the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm. CAMSHIFT's tracking accuracy is compared against a Polhemus tracker. Tolerance to noise, distractors and performance is studied.

*Keywords:* Computer vision, Face tracking, Mean Shift Algorithm, Perceptual User Interface, 3D Graphics Interface

### **25. Paper 29111035: Using RFID to Enhance Mobile Banking Security (pp. 176-182)**

*Zakaria Saleh, MIS Department, IT faculty, Yarmouk University, Irbid, Jordan  
Izzat Alsmadi, CIS Department, IT faculty, Yarmouk University, Irbid, Jordan*

*Abstract* — Mobile banking is introducing a new generation of location-independent financial services using mobile terminals. This facilitates allowing users to make payments, check balances, transfer money between accounts and generate statements of recent transactions on their cellular phones. While providing, anywhere, anytime banking to the user, the service should be secure and security needs to be implemented at various levels, starting from the SIM card security, mobile software security, and secure customer access to banking services. Banks rely on users having their mobile phones with them all the time. Hence, as a mean for security measures, banks can send alerts, anytime, in order to provide an enhanced security and

services. This paper analyzes the security issues in Mobile Banking, and proposes an improved security to the mobile banking services using RFID.

*Key words: Mobile banking, security, RFID, Wireless communication, Pervasive Computing, smart cards, and contactless payment, wireless security, and e-commerce.*

## **26. Paper 30111065: Parcel Management System using GPS Tracking Unit (pp. 183-189)**

*Tahmina Khatoon<sup>1</sup>, Md. Musfique Anwar<sup>2</sup>, Nasrin Sultana Shume<sup>2</sup>, Md. Mizanur Rahman<sup>1</sup>*

<sup>1</sup>Computer Science & Engineering Department, IBAIS University, Dhaka, Bangladesh

<sup>2</sup>Computer Science & Engineering Department, Jahangirnagar University, Dhaka, Bangladesh

*Abstract* - In this paper, the proposed system overcomes the deficiencies of parcel management system for providing parcel information by using operational data extracted from a GPS tracking unit. A GPS tracking unit is a device that uses the Global Positioning System to determine the precise location of a vehicle to which it is attached and to record the position at regular intervals. The recorded location data can be transmitted to a central location database of a remote tracking server using satellite modem embedded in the unit. Tracking server also has satellite modem that receives vehicle location information. This allows the vehicle's location to be displayed against a map backdrop in real-time using customized software to authorized users of the system via website over the internet.

*Keywords: Parcel Management System, GPS, Tracking server, Satellite Modem.*

## **27. Paper 30111081: Generation of Mutation Operators for AOP (pp. 190-194)**

*Mayank Singh, Research Scholar, Uttarakhand Technical University, Dehradun, India*

*Shailendra Mishra, Professor computer science department, KEC dwarahat, Uttarakhand, India*

*Abstract*- Testing of aspect oriented programs is an upcoming challenge for the researchers. Mutation testing has a lot to be undertaken to explore the field of testing of AOP. It is an emerging field of research in testing of aspect oriented programming. Since the effectiveness of mutation testing depends on finding fault types and designing of mutation operators, therefore the effectiveness of testing depends upon the quality of these mutation operators. A detailed study has done on the mutation operators for procedural and object oriented languages, but for aspect oriented language only few researchers had contributed. This paper discusses in detail about the fault types and related mutation operators for AspectJ language. It also proposes the implementation framework of mutation operators automatically.

*Keywords: Mutation Testing, Aspect oriented testing, fault based testing*

## **28. Paper 30111049: Modelling Data Transmission through a Channel Based on Huffman Coding and Encryption Methods (pp. 195-199)**

*Eugène C. Ezin*

*Institut de Mathématiques et de Sciences Physiques, Unité de Recherche en Informatique et Sciences Appliquées, Université d'Abomey-Calavi, République du Bénin*

*Abstract* — Data transmission through a secure channel requires the attention of many researchers. In this paper, on the basis of an alphabet of ciphers and letters, we propose a model for data transmission through a secure channel. This is achieved at two levels. First we associate each distinct symbol with a probability in the message to transmit. By doing so, we modify the well known adaptive Huffman coding method. The obtained alphabet is used to construct the coded message to transmit through a cryptosystem. Therefore, the original message is coded and encrypted before its delivering. The proposed model is examined.

*Keywords-component—Data compression, Huffman coding technique, encryption and decryption algorithms.*

**29. Paper 30111052: Performance Comparison of Multicast Routing Protocols In Ad-Hoc Wireless Networks (pp. 200-202)**

*N. Veena, Dept. of CSE, A N University, Guntur*

*A. VinayaBabu, Dept. of CSE, JNTU College of Engineering, Hyderabad*

*K. Madhukar, Nizam College, Osmania University, Hyderabad*

*B. Sujatha, Dept. of CSE, Nizam College, Hyderabad*

*B. Anandkumar, Dept. of CSE, Rayalasima University, Kurnool.*

*S. Nagaprasad, Dept. of CSE, Acharya Nagarjuna University, Guntur*

**Abstract-** The majority of applications are in the areas where rapid deployment and dynamic reconfiguration are necessary and a wire line network is not available. These include military battlefields, emergency search and rescue sites, classrooms, and conventions, where participants share information dynamically using their mobile devices. Well established routing protocols do exist to offer efficient multicasting service in conventional wired networks. These protocols, having been designed for fixed networks, may fail to keep up with node movements and frequent topology changes in a MANET. Therefore, adapting existing wired multicast protocols as such to a MANET, which completely lacks infrastructure, appear less promising. Providing efficient multicasting over MANET faces many challenges, includes scalability, quality of service, reliable service, security, Address configuration, Applications for multicast over MANET. The existing multicast routing protocol do not addresses these issues effectively over Mobile Adhoc Networks (MANET).

**30. Paper 30111060: A Proposed Ontology Based Architecture to Enrich the Data Semantics Syndicated by RSS Techniques in Egyptian Tax Authority (pp. 203-209)**

*Ibrahim M El-Henawy, Faculty of computer and Informatics, Zagazig University, Zagazig, Egypt*

*Mahmoud M Abd El-latif & Tamer E Amer, Faculty Of Computer and Information, Mansoura University, Mansoura, Egypt*

**Abstract —** RSS (RDF site summary) is a web content format used to provide extensible metadata description and syndication for large sharing, distribution and reuse across various applications; the metadata provided by the RSS could be a bit to describe the web resource; this paper provides a framework for making the RSS not only just for syndicating a little information about news but also for further classification, filtering operations and answering many questions about that news by modeling RSS ontology. The proposed architecture will be applied to handle announcements in the Egyptian Tax authority.

**Keywords-** Semantic Web - RDF – Ontology – RSS – OWL –Protégé - Egyptian Tax

**31. Paper 30111062: Distributed Task Allocation in Multi-Agent System Based on Decision Support Module (pp. 210-215)**

*Sally M. El-Ghamrawy, Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt*

*Ali I. El-Desouky, Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt*

*Ahmed I. Saleh, Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt*

**Abstract-** A Multi-Agent System (MAS) is a branch of distributed artificial intelligence, composed of a number of distributed and autonomous agents. In MAS, an effective coordination is essential for autonomous agents to reach their goals. Any decision based on a foundation of knowledge and reasoning can lead agents into successful cooperation, so to achieve the necessary degree of flexibility in coordination, an agent requires making decisions about when to coordinate and which coordination mechanism to use. The performance of any MAS depends directly with the right decisions that the agents made. Therefore the agents must have the ability of making right decisions. In this paper, we propose a decision support module in a distributed multi-agent system, which enables any agent to make decisions needed for Task allocation

problem; we propose an algorithm for Task Allocation Decision Maker (TADM). Furthermore, a number of experiments were performed to validate the effectiveness of the proposed algorithm (TADM)); we compare the efficiency of our algorithms with recent frameworks. The preliminary results demonstrate the efficiency of our algorithms

*Keywords: Decision Making, Task allocation, Coordination Mechanism, Multi-Agent System (MAS)*

### **32. Paper 30111068: Multiplayer Enhanced Make Square Game in the Net (pp.216-223)**

*Md. Ekramul Hamid, Department of network Engineering, College of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia*

*Sudip Saha, Software Engineer, Protocom Technology, Sydney, NSW-2000 , Australia*

*Mohammed Qayyum, Department of Computer Engineering, College of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia*

**Abstract**— In this work, the authors attempt to create a successful Java socket program to implement the “Make Square” game in the net. The game is very popular among the children. Even though this game is for kids, this one can also be played by an adult because it can be quite tricky, require concentration and a little bit of intelligence. The goal in this game is to make more small squares. A player will win the game, if he can complete maximum number of squares after the game is over. Here client/server technology is used to implement socket programming. Since the game is implemented by java, so it is platform independent and portable. Many players in many different groups can play the game on the net. To make the game more interesting we enhance its feature by adding hidden lines. This makes the game more attractive and challenging. The Java features like Networking, Graphics, Layout Management, Package and Interface, Exception Handling, I/O, Applets, AWT Controls and Event handling etc. [2-4] are used to create the game. The Make Square game consists of more than 1700 lines of code in 12 classes. Five of these classes are part of the server side and rest seven is part of the client side. The Make Square game is running properly in a network.

*Keywords-component; Make Square, node, socket programming, AWT, 3D object, GUI, client/server technology*

### **33. Paper 21111008: Descriptive System for Creating Awareness In The Electoral Process In Nigeria Using Information Technology (pp. 224-229)**

*B. A. Onyekwelu, Department Of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Osun State*

*Dr. D.T. Akomolafe, MBCS, Department Of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Osun State, Nigeria*

**Abstract** – Knowledge is power, as is popularly said, and lack of knowledge of the electoral process of one’s nation makes one a subject, rather than citizen. What makes the difference between citizens and subjects is the type and volume of information possessed. This paper discusses the electoral process in Nigeria in relation to the principal actors in the process, namely, the electorates, the political players, the electoral body, the Judiciary and the Government. They roles of each principal actor are highlighted. The current state of awareness of the electoral process in Nigeria is explained as well as factors leading to this state. Information Technology and its growth in Nigeria are reviewed. The Methodology for creating people’s awareness towards the electoral process in Nigeria is proposed and evaluated. The challenges facing the advancement of Information Technology in the country are enumerated and a conclusion is drawn.

*Keywords: electoral process, information, Nigeria, Government, Technology.*



**34. Paper 24111022: Enhancing and Deriving Actionable Knowledge from Decision Trees (pp. 230-236)**

*P. Senthil Vadivu, HEAD, Department Of Computer Applications, Hindusthan College Of Arts and Science, Coimbatore-641028. Tamil Nadu, India*

*Dr.(Mrs) Vasantha Kalyani David, Associate Professor, Department of Computer Science, Avinashilingam Deemed University, Coimbatore, Tamil Nadu, India*

**Abstract** - Data mining algorithms are used to discover customer models for distribution information, Using customer profiles in customer relationship management (CRM), it has been used in pointing out the customers who are loyal and who are attritors but they require human experts for discovering knowledge manually. Many post processing technique have been introduced that do not suggest action to increase the objective function such as profit. In this paper, a novel algorithm is proposed that suggest actions to change the customer from the undesired status to the desired one. These algorithms can discover cost effective actions to transform customer from undesirable classes to desirable ones. Many tests have been conducted and experimental results have been analyzed in this paper.

**Keywords:** CRM,BSP,ACO, decision trees, attrition

**35. Paper 24111023: Constructing Models for MicroArray Data with Swarm Algorithm (pp. 237-242)**

*Mrs. Aruchamy Rajini, Lecturer in Computer Applications, Hindusthan College of Arts & Science, Coimbatore*

*Dr.(Mrs) Vasantha Kalyani David, Associate Professor, Department of Computer Science, Avinashilingam Deemed University, Coimbatore, Tamil Nadu, India*

**Abstract** - Building a model plays an important role in DNA microarray data. An essential feature of DNA microarray data sets is that the number of input variables (genes) is far greater than the number of samples. As such, most classification schemes employ variable selection or feature selection methods to pre-process DNA microarray data. In this paper Flexible Neural Tree (FNT) model for gene expression profiles classification is done. Based on the predefined instruction/operator sets, a flexible neural tree model can be created and evolved. This framework allows input variables selection, over-layer connections and different activation functions for the various nodes involved. The FNT structure is developed using the Ant Colony Optimization (ACO) and the free parameters embedded in the neural tree are optimized by Particle Swarm Optimization (PSO) algorithm and its enhancement (EPSO). The purpose of this research is to find the model which is an appropriate model for feature selection and tree-based ensemble models that are capable of delivering high performance classification models for microarray data.

**Keywords** --- DNA, FNT, ACO, PSO, EPSO

**36. Paper 24111025: Improved Content Based Image Retrieval Using Color Histogram And Self Organizing Maps (pp. 243-248)**

*R. Priya, HOD I/c, Dept of Computer Science, Sree Narayana Guru College, Coimbatore,India*

*Dr.(Mrs) Vasantha Kalyani David, Associate Professor, Department of Computer Science, Avinashilingam Deemed University, Coimbatore, Tamil Nadu, India*

**Abstract** - Color is a feature of the great majority of content-based image retrieval systems. The conventional color histogram retrieval method is prone to lose the spatial information of colors. This paper proposes two methods; one combines color histograms with spatial information and the second which uses a dimensionality reduction technique that reduces the number of features. The experimental results show that the recall /precision and retrieval time of the proposed method is better than other methods.

**Keywords** – content-based image retrieval, color histogram, spatial information, Self Organizing Map

**37. Paper 30111039: Building an Energy-Efficient Prediction S-MAC Protocol for Wireless Networks (pp. 249-260)**

*Mahmoud Abdel-Aziz El-Sakhawy Othman, Information Technology Sector, Egyptian Financial Supervisory Authority (EFSA).*

*Prof. Dr. Imane Aly Saroit Ismail, Vice Dean for Education & Student Affairs, Faculty of Computers & Information, Cairo University*

**Abstract—** With the rapid development of wireless networking and micro-electro-mechanical systems (MEMS), wireless sensor networks (WSNs) have been immersed. WSNs consist of large amount of small, low-end, resource constrained devices, called sensors. Since sensor nodes are usually intended to be deployed in unattended or even hostile environments, it is almost impossible to recharge or replace their batteries. One of the most important research issues in the wireless sensor networks is to extend the network lifetime by energy efficient battery management. So, there are a lot of approaches that are designed to reduce the power consumption of the wireless sensor nodes. In this paper; a new protocol named "prediction S-MAC protocol" is proposed to reduce the power consumption of the wireless sensor nodes and to improve their performance compared to the previous S-MAC protocols.

**Keywords -** *Wireless sensor network; Sensor medium access control (S-MAC) protocol; periodic listen and sleep; adaptive listen, prolong listen, prediction S-MAC protocol.*

**38. Paper 30111066: Simulation of Grover's Algorithm Quantum Search in a Classical Computer (pp. 261-269)**

*A.B. Mutiara, & R. Refianti*

*Faculty of Computer Science and Information Technology, Gunadarma University, Jl. Margonda Raya No.100, Depok 16464, Indonesia*

**Abstract—** The rapid progress of computer science has been accompanied by a corresponding evolution of computation, from classical computation to quantum computation. As quantum computing is on its way to becoming an established discipline of computing science, much effort is being put into the development of new quantum algorithms. One of quantum algorithms is Grover's algorithm, which is used for searching an element in an unstructured list of N elements with quadratic speed-up over classical algorithms. In this work, Quantum Computer Language (QCL) is used to make a Grover's quantum search simulation in a classical computer document.

**Keywords:** *Grover's Algorithm, Quantum Computer Language, Hadamard-Transform*

**39. Paper 30111071: Implementation of a new Fuzzy Based Load Balancing Algorithm for Hypercubes (pp. 270-274)**

*Sameena Naaz, & Afshar Alam, Department of Computer Science, Jamia Hamdard, Hamdard University New Delhi, India*

*Ranjit Biswas, Institute of Technology and Management, Gurgaon, Haryana*

**Abstract —** Distributed computing systems are becoming increasingly available because of the rapid decrease in hardware cost and the advances in computer networking technologies.. It is frequently observed that in a computing environment with a number of hosts connected by a network, the hosts are often loaded differently. In typical distributed system task arrive at the different nodes in random fashion. This causes a situation of non-uniform load across the different nodes. Load imbalance is observed by existence of nodes that are highly loaded while the others are lightly loaded or even idle. Such situation is harmful to the system performance in terms of response time and resource utilization. In the work presented in this paper we have tried to analyze the effect of using fuzzy logic to deal with the problem of load balancing in hypercube model.

**Keywords -** *Load Balancing, Fuzzy Logic, Hypercubes, Response Time*

**40. Paper 23111017: Towards a better assessment of the durations in PERT Method (pp. 275-281)**

*Nasser Eddine MOUHOU, Computer science department, Sétif University, Algeria*  
*Hocine BELOUADAH, Computer science department, M'sila University, Algeria*  
*Abdelhak BOUBETRA, Computer science department, Bordj Bou Arréridj University, Algeria*

**Abstract**— During many years, two of the most popular approaches to the project management were used. It is about the critical path method (CPM) and the PERT method (Program and Evaluation Review Technique). They were based on modeling by digraphs. CPM is unaware of the stochastic nature of the activities and brings back the model to a deterministic model. PERT holds in account this case but the estimation of the activities is despoiled with several errors. In this paper, this technique is presented. It will be followed by an analysis, criticisms and new proposals to make corrections to this method.

**Keywords** - Critical Path Method (CPM), PERT method, stochastic PERT

**41. Paper 30111041: Network Anomaly Detection and Visualization using Combined PCA and Adaptive Filtering (pp. 282-284)**

*Altyeb Altaher & Sureswaran Ramadass*  
*NAv6 Center of Excellence, Universiti Sains Malaysia USM, 11800 Penang, Malaysia*  
*Noureldeen Abdelrahman & Ahmed Khalid*  
*Faculty of Computer Sciences and IT, University of Sciences and Technology, Sudan*

**Abstract** - In recent years network anomaly detection has become an important area for both commercial interests as well as academic research. This paper provides a Combined Principal Component Analysis (PCA) and Filtering Technique for efficient and effective detection and identification of network anomalies. The proposed technique consists of two stages to detect anomalies with high accuracy. First, we apply the Principal Components Analysis to transform the data to a new coordinate system such that the projection on the coordinate contains the greatest variance. Second, we filter traffic to separate between the normal and anomalous traffic using adaptive threshold. Our analysis results from network-wide traffic datasets show that our proposed provides high detection rate, with the added advantage of lower complexity

**Keywords**- Network anomaly detection, principal component analysis , network anomaly visualization, adaptive network traffic filter.

**42. Paper 30111045: A Two Dimensional Approach to Back and Forth Software Process Model (pp. 285-291)**

*Sameera Amjad, Wasi Haider Butt, Farooque Azam*  
*Department of Computer Engineering, College of Electrical & Mechanical Engineering, National University of Science & Technology (NUST), Rawalpindi, Islamabad*

**Abstract**— Many software development process models have been documented but none of them gives a detailed methodology for change management. This article proposes a novel software development process model which realizes the inherent nature of requirement changes and provides a methodology to accommodate these changes. A detailed literature survey was conducted to explain the difference of the proposed model with existing software development approaches. The proposed novel model namely, the Back and Forth software process model uses two methods to present the development methodology.

**Keywords**- Software Development Life Cycle; Software Process Models



**43. Paper 30111051: Performance Comparison of Block Truncation Coding based Image Retrieval Techniques using Assorted Color Spaces (pp. 292-297)**

*Dr. H.B. Kekre, Senior Professor, Computer Engg. Dept., MPSTME, SVKM's NMIMS University, Mumbai, India*

*Sudeep D.Thepade, Ph.D. Research Scholar & Associate Professor, Computer Engg. Dept. MPSTME, SVKM's NMIMS University, Mumbai, India*

*Varun Lodha, Pooja Luthra, Ajoy Joseph, Chitrangada Nemani*

*B.Tech (IT) Students, MPSTME, SVKM's NMIMS University, Mumbai, India*

**Abstract**— The paper presents exhaustive performance comparison of image retrieval techniques based on block truncation coding (BTC) using assorted color spaces. Including RGB total ten color spaces are considered for applying BTC to extract the feature vector in CBIR techniques. Further the image tiling is added to get three CBIR techniques per color space. In all performance comparison of thirty image retrieval techniques is done with the help of generic image database having 1000 images spread across 11 categories. For each proposed CBIR technique 55 queries (5 per category) are fired on the generic image database. To compare the performance of image retrieval techniques average precision and recall are computed of all queries. The results have shown the performance improvement (higher precision and recall values) with proposed color-BTC methods compared to gray-BTC in all color spaces except 'rgb' color space. Image tiling does not help to improve the performance in the chromaticity-luminance based color spaces (Kekre's LUV, YCbCr, YUV, YIQ, Kekre's YCgCb), while it helps in non-luminance color spaces (RGB, HSV, XYZ, HIS). Overall Kekre's LUV color space based BTC gives best performance in image retrieval.

**Keywords** — CBIR, BTC, Color Space, Image Tiling, VQ, RGB, HSV, XYZ, HIS, rgb, Kekre's LUV, YCbCr, YUV, YIQ, Kekre's YCgCb

**44. Paper 30111055: Development of a Project-Based Learning Approach in Requirement Engineering (pp. 298-303)**

*Nor Azliana Akmal Jamaludin, Fakulti Industri Teknologi Maklumat, Universiti Industri Selangor, Kampus Bestari Jaya, Jalan Timur Tambahan, 45600 Bestari Jaya, Selangor, Malaysia.*

*Shamsul Sahibuddin, Advanced Informatics School, UTM International Campus, Universiti Teknologi Malaysia, Malaysia Jalan Semarak, 54100 Kuala Lumpur Malaysia*

*Kamaruzaman Jusoff (Corresponding author) Faculty of Forestry, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia*

*Nur Hafizah Hidayat, Fakulti Industri Teknologi Maklumat, Universiti Industri Selangor, Kampus Bestari Jaya, Jalan Timur Tambahan, 45600 Bestari Jaya, Selangor, Malaysia.*

**Abstract**— Project failure is due to the lack of Requirement Engineering (RE) practice. The Industry needs to allocate another cost to send their employee for additional training before the employee can contribute to the job specification. It indicates that current practices of delivery mechanisms at the university fail to deliver graduates with employability skills. The objective of this research is to identify weaknesses in current practice of teaching Software Engineering undergraduate in Requirement Engineering. Additionally, this paper emphasized that Project-Based Learning (PjBL) is a right method for delivery mechanisms to enhance Software Engineering undergraduate skills particularly in RE. The PjBL is a superset to Problem-Based Learning, Individual-Collaborative Learning and Product-Based Learning. The intersection can strongly assist in the learning environment. Future work should be carried out to design the framework of PjBL, measuring the effectiveness of PjBL and the electronic Learning eNvironment (eLIN) system as a supportive tools to make PjBL successful.

**Keywords**- Software Engineering education; Project-Based Learning (PjBL); Requirement Engineering; Problem-Based Learning; Individual & Collaborative Problem Solving and Product-Based Learning.

**45. Paper 30111061: Adaptation of GQM Method for Evaluating the Performance of Software Project Manager (pp. 304-307)**

*Nagy Ramadan Darwish, Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University Cairo, Egypt*

*Mahmoud M. Abd Ellatif, Department of Information Systems, Faculty of Computers and Information, Mansoura University Mansoura, Egypt*

**Abstract**—This paper is concerned with evaluating the performance of software project manager using Goal Question Metrics (GQM) method. It clarifies the Software Project Management (SPM) domains and the performance metrics of each domain. This paper presents the basic concepts of GQM method. Based on a combination of statistical techniques, this paper presents how to apply GQM method to evaluate the performance of a software project manager. A software company can use the proposed approach to track, evaluate, control, correct, and enhance the performance of software project managers to increase the success rate of software projects.

**Keywords:** *Software ;Project Manager; Performance; Evaluation - GQM – Metrics – Performance Report*

**46. Paper 30111069: Adaptive E-Learning System based on Semantic Web and Fuzzy Clustering (pp. 308-315)**

*Khaled M. Fouad, Computer Science Dep. - Community College - Taif University, Kingdom of Saudi Arabia (KSA)*

*Mofreh A. Hogo, Computer Science Dep. – Computers and Information Systems College - Taif University, Kingdom of Saudi Arabia (KSA).*

*Shehab Gamalel-Din, Computers Dept, Faculty of Science, King Abdul-Aziz University, Kingdom of Saudi Arabia (KSA)*

*Nagdy M. Nagdy, Systems and Computers Engineering Dep. – Faculty of Engineering – AlAzhar University, Egypt.*

**Abstract**— This work aims at developing an adaptive e-learning system with high performance to reduce the challenges faces elearners, the instructors and provides a good monitoring system for the complete e-learning systems as well as the system structure. The work presents the different phases for the system development of the adaptive system as: the first stage is the collection of the e-learners documents, the second stag is the documents representation including the frequency count and the weighting of the documents with its frequencies, the third stage is the prediction and clustering of e-learners interests using the fuzzy clustering method and the statistical K-means clustering method. The results obtained from this work shows that we have to have different e-learners ontologies using the results of the clustering methods which reflect the e-learners interests. Finally the work concluded the suggestions as well as the recommendations for the instructors and the systems administrators.

**Keywords-***component; E-Learning; Semantic Web; Fuzzy Clustering; User model; User Model Representation*

**47. Paper 30111072: An Innovated Server-Based Desktop Sharing Platform (pp. 316-324)**

*Ayad H. Abdulqader, Salah Noori Saleh, Omar Amer Abouabdalla, Sureswaran Ramadas, National Advance IPv6 Centre of Excellence (NAv6), USM, 11800 Penang, Malaysia*

*Hasan S. M. Al-Khaffaf, School of Computer Science, USM, 11800 Penang, Malaysia*

**Abstract** - In this paper, a server-based Desktop Sharing Platform (DSP) is proposed. The proposed platform is designed to work with any direct-connection Remote Desktop System (RDS) without modifying or adding any extra models to those systems for both server's and clients' sides. The existing RDS systems' limitations in terms of bandwidth consumption, collaboration session initiation, and connectivity issues will be overcome by adopting the proposed platform. The proposed platform is easily adapted to work with any direct-connection RDS system. Incorporating the proposed platform will improve the performance and

efficiency of existing RDS systems. As a result, better utilization of computer system resources in terms of bandwidth and processing power is achieved by minimizing the data transfer and processing power from n users to only one user.

*Keywords-* Computer Supported Cooperative Work; Remote Display System; Thin-Client Computing

**48. Paper 30111073: Design of Hybrid Ontologies for Mediation System Applied to the E-learning Platform (pp. 325-329)**

*Otman ABDOUN, Laboratory LIST, ERIT, FST – Tangier, Tangier, Morocco*

*Jaber EL BOUHDIDI, Laboratory LIST, FST – Tangier, Tangier, Morocco*

*Mohamed GHAILANI, Laboratory LIST, FST – Tangier, Tangier, Morocco*

*Abdelhadi FENNAN, Laboratory LIST, FST – Tangier, Tangier, Morocco*

*Abstract*—This work falls within the scope of E-learning is important for several reasons. First, resources are structured (educational needs) and therefore easier to annotate. Second, there is a curriculum (or Education Plan) that ensures the semantic integration of resources. Third, services are available to the teacher and learner. And finally, post evaluation of knowledge acquired by the learner, to verify the adequacy of resources presented to the learner, and indirectly the appropriateness of teaching strategies implemented to follow up resources and services. First of all, it describes the problems of integrating multiple sources of educational and placed in the ontology integration process, then treated mediation services, and their contribution on an E-learning platform.

*Keywords-* E-learning; Mediation Services; Hybrid Ontologies

**49. Paper 30111074: Extracting Membership Functions Using ACS Method via Multiple Minimum Supports (pp. 330-336)**

*Ehsan Vejdani Mahmoudi, Islamic Azad University, Mashhad Branch, Young Researchers Club, Mashhad, Iran*

*Masood Niazi Torshiz, Department of computer Engineering, Islamic Azad University - Mashhad Branch, Mashhad, Iran*

*Mehrdad Jalali, Department of computer Engineering, Islamic Azad University - Mashhad Branch, Mashhad, Iran*

*Abstract*— Ant Colony Systems (ACS) have been successfully applied to different optimization issues in recent years. However, only few works have been done by employing ACS method to data mining. This paper addresses the lack of investigations on this study by proposing an ACS -based algorithm to extract membership functions in fuzzy data mining. In this paper, the membership functions were encoded into binary bits, and then they have given to the ACS method to discover the optimum set of membership functions. By considering this approach, a comprehensive exploration can be executed to implement the system automation. Therefore, it is a new frontier, since the proposed model does not require any user-specified threshold of minimum support. Hence, we evaluated our approach experimentally and could reveal this approach by significant improving of membership functions.

*Keywords-* fuzzy data mining; multiple minimum supports; association rule; membership functions; ant colony system.

**50. Paper 31101085: Enhancing K-Means Algorithm with Semi-Unsupervised Centroid Selection Method (pp. 337-343)**

*R. Shanmugasundaram and Dr. S. Sukumaran*

*Abstract*— The k-means algorithm is one of the frequently used clustering methods in data mining, due to its performance in clustering massive data sets. The final clustering result of the kmeans clustering algorithm is based on the correctness of the initial centroids, which are selected randomly. The original k-

means algorithm converges to local minimum, not the global optimum. The k-means clustering performance can be enhanced if the initial cluster centers are found. To find the initial cluster centers a series of procedure is performed. Data in a cell is partitioned using a cutting plane that divides cell in two smaller cells. The plane is perpendicular to the data axis with very high variance and is intended to minimize the sum squared errors of the two cells as much as possible, while at the same time keep the two cells far apart as possible. Cells are partitioned one at a time until the number of cells equals to the predefined number of clusters, K. The centers of the K cells become the initial cluster centers for K-means. In this paper, an efficient method for computing initial centroids is proposed. A Semi Unsupervised Centroid Selection Method is used to compute the initial centroids. Gene dataset is used to experiment the proposed approach of data clustering using initial centroids. The experimental results illustrate that the proposed method is very much apt for the gene clustering applications.

*Index Terms*— Clustering algorithm, K-means algorithm, Data partitioning, initial cluster centers, semi-supervised gene selection

#### **51. Paper 31101086: A Survey on Static Power Optimization in VLSI (pp. 344-349)**

*A. Janaki Rani and Dr. S. Malarkkan*

*Abstract*---Power has become one of the primary constraints for both the high performance and portable system design. The growing market of battery powered electronic systems like cellular phones, personal digital assistants demands the design of microelectronic circuits with low power consumption. Power dissipation in these systems may be divided into two major components namely static and dynamic power dissipation. The static power is the standby power that is wasted even if the device is not performing any function. As technology scales down the static power dissipation is dominant in VLSI circuits which are mainly due to leakage current in transistors. Hence a focus is necessary on the leakage currents. These leakage currents are mainly due to sub-threshold leakage and gate oxide leakage. The sub-threshold leakage is dominant which can be minimized by reducing the supply voltage, reducing the transistor size, decreasing the temperature and increasing the threshold voltage. In this paper a survey is presented on static power optimization in VLSI. It presents the possible solutions to reduce the leakage power in various digital logic circuits like CMOS, I<sup>2</sup>C etc.

*Index Terms*—Leakage, Low-Power, Power Gating, Semicustom, Input Vector Control, Body Bias Control, Sleep Transistor Sizing, Sleepy Stack, Zigzag Power Gating (ZPG)

#### **52. Paper 31101087: A Survey on Minimizing Energy Consumption of VLSI Processors Using Multiple Supply Voltages (pp. 350-355)**

*B. Sathiyabama and Dr. S. Malarkkan*

*Abstract*—Due to the continuous increase in earth's population, adequate supply of resources is going to be a major issue. One basic essential resource in rising demand is energy and in particular electrical energy. The contributions of the scientific community toward the goal of sustainability with regard to energy consumption of embedded systems are previously discussed in many research works. Low power has become one of the major design issues due to the increased demand in personal computing devices and portable communication system. In this paper a survey on minimizing energy consumption of VLSI Processors using multiple supply voltages is presented. This survey discusses on search method for a scheduling and module selection problem using multiple supply voltages so as to minimize dynamic energy consumption under time and area constraints. The algorithm based on a genetic algorithm is surveyed to find near-optimal solutions in a short time for large-size problems. The literature related to the multiple supply voltages with genetic approach and energy consumption minimization in various VLSI systems is presented.

*Keywords*— Energy minimization, Functional pipelining, Multiple supply voltages, dynamic power, scheduling



**53. Paper 26111028: System Integration for Smart Paperless Ship (pp. 356-364)**

*O. Sulaiman, A.H. Saharuddin, Faculty of Maritime Studies and Marine Science, Univeristy Malaysia Terengganu, Kuala Terengganu, Malaysia*

*A.B.S. Kader, Faculty of Mechanical -Marine Technology, University Technology Malaysia, Kuala Terengganu, Malaysia*

**Abstract**— Sea Transportation provides a safe and reliable source for moving people and cargo across the ocean. The government and private sector provide these services, while the Government moves personnel and cargo to support global peacekeeping activities while the civilian sea transportation activities focus on providing leisure cruises and the shipment of consumer goods. These safe and efficient services are obtained through the cooperative efforts of the government and civilian sea carriers, and seaports throughout the world required connectivity ranging from within ship system integration and ship shore operation, which has been much facilitated by evolution of computer age. The use of the use of new information technology and interfacing all the associated service areas of maritime industry- sea and shore has lead to reducing papers and hence cutting of threes and beneficial environmental benefit of excess water absorption and greater capture of carbon dioxide. Human race has achieved much civilization and development in recent years until it seem as development is closed to the peak. However, new philosophy under are being promoted in recent years include proactive behaviors, recycling, system integration and conservation to make all what has been built meaningful and efficient. This paper discuss how system integration under smart ship concept within ship and shore.

**Keywords**- sytem integration, paperless, ship, electronics waste



# Chaotic hash function based on MD5 and SHA-1 hash algorithms

Mouad HAMRI <sup>#1</sup>, Jilali Mikram <sup>#2</sup>, Fouad Zinoun <sup>&3</sup>

<sup>#</sup> *Mathematics and computer science department, Science University of Rabat-Agdal*

*4 Avenue Ibn Battouta Rabat Morocco*

<sup>&</sup> *Economical sciences and management department, University of Meknes*

<sup>1</sup> hamri.mouad@gmail.com

<sup>2</sup> mikram@fsr.ac.ma

<sup>3</sup> fouad.zinoun@gmail.com

**Abstract**—In this work an implementation of a 128 bits chaotic hash function is presented. We have used the logistic map as the chaotic dynamical system of our algorithm and for the initial condition generation of the chaotic map, we've used the two hash functions MD5 and SHA-1. The main idea is to use the MD5 and SHA-1 digits of the a message to construct the initial condition of the logistic map and using this map we will generate the digit of this message. The algorithm was implemented using JAVA programming language and its performance was studied by comparing it to MD5 hash function.

**Keywords:** Cryptography, Hash functions, MD5, SHA, Dynamical systems, Chaos theory.

## I. INTRODUCTION

Hash functions [14] play an important role in the cryptography domain with many applications as digital signatures, message authentication codes and many other authentication forms. A hash function can be simply defined as a function that convert a message of large size to a digit with a fixed and smaller size and a more precise definition can be announced as following:

**Hash function:** A hash function is a function  $h$  which has, as a minimum, the following two properties:

- 1) *compression*:  $h$  maps an input  $x$  of arbitrary finite bitlength, to an output  $h(x)$  of fixed bitlength  $n$ .
- 2) *ease of computation*: given  $h$  and an input  $x$ ,  $h(x)$  is easy to compute.

An ideal cryptographic hash function must have the following properties:

- It's easy to compute the digit of any given message.
- From a given digit, it's infeasible to find a corresponding message.
- It's also infeasible to find two different messages with the same digit.

Because this kind of function can't be a bijection, many messages have the same digit so the real challenge is to construct a hash function in a way that it is very hard and costly to find a collision (two messages with the same digit). Many algorithms have been already developed to construct hash function as MD5 and SHA-1 algorithms but their security

level was reduced because of the number of attacks and collisions found against them. The chaotic dynamical systems have become since the last two decades a very powerful tool used in the modern cryptography, their high sensitivity to initial condition and the ergodicity property guarantee a high level of randomness evolution while they're deterministic. Our main idea is to use a chaotic dynamical system for the digit generation, and we use both MD5 and SHA-1 algorithms to generate the initial condition for each message. We'll start with an introduction to chaotic dynamical systems and to logistic map then we'll present our algorithm with the main results and performance tests comparing it to MD5 algorithm.

## II. CHAOTIC DYNAMICAL SYSTEMS AND LOGISTIC MAP

A dynamical system ([1-3],[11-12]) can be defined as a system of equations describing the evolution of a mathematical model where the model is fully determined by a set of variables.

One of the most famous discrete dynamical system is the logistic map (that will be used in our algorithm) defined on the set  $[0, 1]$ . The logistic map is written:

$$x_{n+1} = rx_n(1 - x_n)$$

Where  $x_0$  represent the initial condition,  $n \in \mathbb{N}$  and  $r$  is positive real number.

In reality, there is no universal definition for chaotic dynamical systems. The following definition tries to define a chaotic dynamical system using three ingredients that almost everyone would agree on.

**Chaotic dynamical system:** Let  $f : X \rightarrow Y$  a function ( $X, Y \subseteq \mathbb{R}$ ).

The dynamical system  $\dot{x} = f(x)$  is said to be **chaotic** if the following proprieties are satisfied:

1- **Sensitive dependance on initial conditions:**  $\forall \beta > 0, \exists \varepsilon > 0$  there exists a point  $y_0 \in X$  and  $k > 0$ , such that:  $|x_0 - y_0| < \beta \Rightarrow |x_k - y_k| > \varepsilon$ .

2- **Density of periodic orbits:** The ensemble of periodic orbits:  $\{x_0 \in X, \exists k > 0, x_k = x_0\}$  is dense in  $X$ .

3- **Deterministic:** means that the system has no random or

noisy inputs or parameters.

The definition above is applied to both discrete and continuous dynamical systems.

The logistic map is a chaotic dynamical system and presents a very high sensitivity to initial conditions for  $r$  between about 3.57 and 4 (approximatively).

The below figure shows the bifurcation diagram of the logistic map:

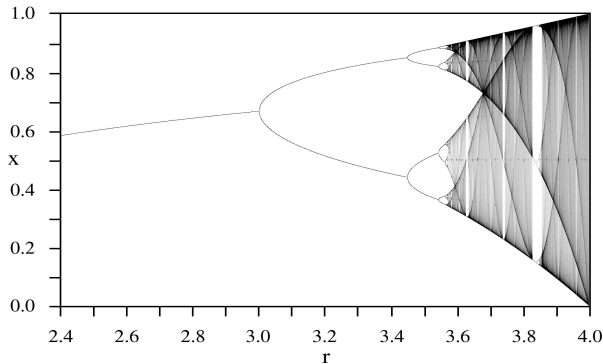


Fig. 1. Bifurcation diagram of the logistic map

### III. THE ALGORITHM

The proposed hash function (CHF-MD5-SHA1) takes as input a message  $M$  and return its digit  $H(M)$ .

The logistic map is used to generate  $H(M)$  by using the values of the function starting from an initial condition obtained using the MD5 and SHA-1 digits of the message  $M$ .

To present the algorithm, we present the following notations:

$M$	The message input
$MD5(M)$	$M$ digit using MD5 hash function
$SHA32(M)$	First 32 bytes of $M$ digit using SHA-1 hash function
$X \text{ xor } Y$	$X$ and $Y$ xored
$L(x_0, N)$	Value of the logistic map starting from $x_0$ after $N$ iterations
$F$	A map from the set of 32 bytes numbers to the interval $[0, 1]$ $F(X) = X \times 2^{-128}$
$H(M)[i]$	$i$ th byte of the digit $H(M)$

The algorithm description can be summarized as following:

- **Begin:**

- Step 1: Let  $IC$  be defined as:  
 $IC = MD5(M) \text{ xor } SHA32(M)$ .
- Step 2: Define the initial condition of the logistic map as:  
 $x_0 = F(IC)$ .
- Step 3: Run the logistic map and stop after  $N$  iterations.
- Step 4: For  $i$  from 1 to 32 do:
- $H(M)[i] = \text{Round}(L(x_0, N + i) * 1000) \bmod 16$

- **End**

In the first step, we generate the initial condition by using the MD5 and SHA-1 digits of the message  $M$ .

As the initial condition should be a real number from the interval  $[0, 1]$ , we use the map  $F$  to convert this 32 bytes number to a number from the desired interval.

In step three, we run the logistic map by  $N$  iterations using as initial condition, the initial condition calculated in the previous step. The aim of this step is to let the system evolve so that all the chaotic properties will have their affect.

In the last step, we simply construct the digit  $H$  byte by byte using the 32 values of the logistic map after the  $N$  iterations. The fact that we're using both MD5 and SHA-1 algorithms to generates the initial condition will make the algorithm more secure as in addition to find  $x_0$  knowing only  $L(x_0, N)$  (which is infeasible for chaotic dynamical systems), we need also to find  $M$  knowing  $(MD5(M) \text{ xor } SHA1(M))$  which is equivalent to break MD5 and SHA1 at the same time.

### IV. RESULTS AND PERFORMANCE TESTS

In this section we will present some results and performance tests of the implemented algorithm comparing it with MD5 hash function.

The computation was done using a PC with the following characteristics: 1,8GHz Core(TM) 2 Duo, 1.00 Go RAM and 120 Go hard-disk capacity.

#### A. Numerical computations

We present in the table below, the results of numerical computations where we calculate for a sample of messages, their digits using the implemented algorithm and MD5 hash function.

For the parameter  $r$  of the chaotic logistic map we use the value:  $r = 4$ .

Hash Function	Message	Digit
MD5	a	0cc175b9c0f1b6a831c399e269772661
	Digit(m)= DigiT(m)=	7cc5d72b3f495cd5aa9321b2c4325769 ac93273fa7b6170704094f3760f5ceb0
CHF-MD5-SHA1	a	46fce5ba0a27ba586491e375723278a3
	Digit(m)= DigiT(m)=	21dace9770da1d1b979bea82cae92fb0 45fdc693875d5936d96328bf78dcf4f7

It's clear from the example above that a small change in the message will lead to a huge difference in the digit for both algorithms.

#### B. Algorithm Security

In this part we're using as security measure, the number of operations needed to find a collision using a brute force attack. The below table summarizes the security measure results:

Hash Function	Digit size(bits)	Brute force attack
MD5	128	$2^{30}$ [15]
CHF-MD5-SHA1	128	$2^{64}$

From this analysis we can see clearly that the proposed algorithm is more secure than MD5 algorithm.



### C. Randomness test

To prove that CHF-MD5-SHA1 produce the digits in a randomly manner, we have performed a randomness test. The test consist of taking different sizes of messages (we've chosen the sizes: 1KB, 5KB, 10KB, 100KB and 1MB) and from each size we construct a set of 1000 messages. The first message is the message with all bits equal to 1 then to have the second message we change one bit to 0, to have the third message we change another bit to 0 in the second message so that the third message contains two 0 bits. The message number  $i$  will contain  $i - 1$  bits equal to 0.

The table below presents the set of message for a given size prepared for the randomness test:

Message number	Message bits
1	1...1111111
2	1...1111110
3	1...1111100
$\vdots$	$\vdots$
$i$	1..10..0000
$\vdots$	$\vdots$

After building the lists of 1000 messages for all the sizes, we calculate their digit using CHF-MD5-SHA1 and MD5 then we compute for every two consecutive digits  $D_k$  and  $D_{k+1}$  (digits of two consecutive messages) the rate of change defined as:

$$RC = \sum_{i=1}^{32} \delta_{D_k[i], D_{k+1}[i]}$$

Where  $\delta$  is the Kronecker symbol and  $D_k[i]$  is the  $i$ th byte of the digit  $k$ .

We compute then the average and standard deviation for every size. The table below presents the different results of the two hash functions:

Messages size	CHF-MD5-SHA1		MD5	
	Average	StDev	Average	StDev
1KB	29,781	2,372	29,859	2,343
5KB	29,835	2,389	29,821	2,348
10KB	29,933	2,347	29,855	2,323
100KB	29,834	2,334	29,820	2,360
1MB	29,911	1,305	29,897	2,321

From the randomness test result, it's clear that except the case where the messages size was 1KB, CHF-MD5-SHA1 has a higher level of randomness. After this result we can say that CHF-MD5-SHA1 hash function is better than MD5 algorithm as it's more secure and has a higher level of randomness.

### V. CONCLUSION

In this paper we have constructed a hash function (CHF-MD5-SHA1) using a chaotic dynamical system and using also the famous hash function MD5 and SHA-1. The performance of this function was tested against MD5 hash function and we've concluded from these tests that the developed algorithm is more secure than MD5 and has a better level of randomness

as well which means that it's practically better than MD5. Chaotic dynamical systems are a great tool that can help to conceive very powerful cryptosystems and hash functions. Some cryptanalysis attacks should also be developed as well to increase the performance of such algorithms.

### REFERENCES

- [1] V.I. Arnold, *Ordinary Differential Equations* (Springer, Berlin, 1977, 1992).
- [2] James D. Meiss, *Differential Dynamical Systems* (SIAM 2007).
- [3] V.I. Arnold, *Geometrical Methods in the Theory of Differential Equations* (Springer, Berlin, 1988).
- [4] V.I. Arnold and D.V. Anosov, *Dynamical Systems I* (Springer, Berlin, 1988).
- [5] Yu. A. Mitropolsky and A.K. Lopatin, *Nonlinear Mechanics, Groups and Symmetry* (Kluwer, Dordrecht, 1995).
- [6] Yu. N. Bibikov, *Local Theory of Analytic Ordinary Differential Equations, Lecture Notes in Mathematics* (Springer, Berlin, 1979).
- [7] A.D. Bruno, *Local Methods in Nonlinear Differential Equations* (Springer, Berlin, 1989).
- [8] V.I. Arnold, *Mathematical methods of classical mechanics*, (2nd edition, Springer, 1989).
- [9] Bellman R., *Stability theory of differential equations*, (MGH, 1953).
- [10] R.C. Schmitt and K. Thompson, *Nonlinear Analysis and Differential Equations, An Introduction* (Aug 2000).
- [11] P. Hartman, *Ordinary Differential Equations* (John Wiley and Sons 1964)
- [12] George D. Birkhoff, *Dynamical systems* (American Mathematical Society 1991)
- [13] Floriane Anstett, *Les systemes dynamiques chaotiques pour le chiffrement : synthese et cryptanalyse (These)* (Universite Henri Poincare - Nancy 1)
- [14] A. Menezes, P. van Oorschot, and S. Vanstone *Handbook of Applied Cryptography* (CRC Press, 1996)
- [15] J. Black, M. Cochran and T. Highland *A Study of the MD5 Attacks: Insights and Improvements* (University of Colorado at Boulder, University of Texas at Austin)

# Concept Measure Supported Aesthetics Ontology Construction for Videos

Dr Sunitha Abburu

Professor & Director, Department of Computer Applications  
Adhiyamaan College of Engineering  
Hosur, pin-635109, Tamilnadu, India.  
ausunithaa@yahoo.co.in

**Abstract**—Entertainment plays a vital role in human life. Multimedia conquers top position in the entertainment world. Video stands first among the multimedia entertainment. The rapid growth of the videos has resulted in the emergence of numerous multimedia repositories that require efficient and effective video storage, semantic annotation, indexing and retrieval systems. The introduction of ontologies in multimedia retrieval system can improve the precision and recall rate effectively. The performance of the annotation and the retrieval system increases with the support of the domain ontology. Most of the video annotation, indexing and the retrieval systems focus on the semantic concepts like objects, people, location, events, actions etc. But most of the multimedia systems are filled with human and their emotions. Any multimedia system like cinema, news videos, sports videos, and any domestic functional videos tries to capture the emotions of the human involved in the occasion. A video retrieval system will be complete if the system identifies, captures and represents the emotions of the humans. This paper focus on identification and representation of human emotions and the intensity of the emotions are represented using the fuzzy logic. The concept of Navarasra has been brought in to video retrieval system to classify the human emotions. The proposed approach is generic and flexible. It is designed and constructed for all videos where human beings and their emotions are been captured. A practical implementation is done using Protégé as an Ontology developing tool.

**Keywords**—component; Video Semantics, Concept measures, Ontology, Retrieval, Human Emotions.

## I. INTRODUCTION

Entertainment plays a vital role in the human life. Multimedia plays an important role in the entertainment world. Video systems stand first in multimedia presentations. The reason why a lot of research work, in the area of multimedia, is carried out on video data compared with other multimedia data types is twofold. First, video contains audio, visual, text information etc. It is the most powerful and at the same time most complex, voluminous, unformatted, and unstructured of all media used for conveying information. Hence, representing video information to enable effective and efficient retrieval is an interesting problem. Second, news videos, cinema videos and the recent trend of capturing any occasion, function, event in videos, raises the demand for the efficient video storage, semantic annotation, indexing and retrieval systems. Video retrieval systems which focus on the low level features and ignore the semantic are less efficient and effective. In order to

improve the effectiveness and the efficiency of the video retrieval system video semantics should be identified, represented and must be used during video object retrieval. The semantic annotation generation can be manual or automated. The annotation generation system and the retrieval system performance increases by considering the ontology. Ontology plays a vital role in artificial intelligence, semantic web, software engineering, information retrieval, knowledge representation, knowledge sharing, knowledge integration, knowledge reuse, and so on. It is a well known fact that the performance of the annotation and the retrieval system increases with the support of the domain ontology. The introduction of ontologies in multimedia retrieval system can improve the precision and recall rate effectively. Focusing on the completeness and the effectiveness of the video retrieval system raises the need for multiple sub ontologies by considering the various aspects of the video semantics. The literature shows that most of the video annotation, indexing and the retrieval systems focus on the semantic concepts like objects, people, location, events, actions etc. But most of the videos are filled with human and their emotions. Any multimedia system like cinema, news, sports, and any domestic functional videos tries to capture the emotions of the human involved in the occasion. A video retrieval system will be complete if the system identifies, captures and represents the emotions of the humans and the intensity of the emotions. Intense understanding of the user's perspective and their expectations towards the video retrieval system is essential. The user queries can be pertaining to the human emotions and the intensity of the emotions involved in the video. This paper focus on identification and representation of human emotions and the intensity of the emotions are represented using the fuzzy logic.

The rest of the paper is organized as follows. Literature survey report is in section 2. Section 3 discusses the proposed method for identification and representation of human emotions and their intensities. In section 4, we present a practical implementation and experimental results on aesthetic ontology construction. Finally, we conclude with a summary and future work in section 5.

## II. RELATED WORK

What is an ontology, the answer is twofold as given in [1], philosophical and computing. In the context of philosophy, Ontology is the philosophical study of the nature of being,

existence or reality, as well as the basic categories of being and their relations. Traditionally listed as a part of the major branch of philosophy known as metaphysics, ontology deals with questions concerning whether entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences. In the context of computer and information sciences [2], an ontology is an explicit specification of a conceptualization. An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Ontology is a kind of concept model that could describe system at the level of semantic knowledge. It aims to access knowledge in a domain in a general way and provides a common understanding for concepts in the domain so as to realize knowledge sharing and reuse among different application programs and organizations. As a new kind of knowledge organization tool and an ontological commitment is an agreement to use a defined vocabulary by a group of people agreed upon in a coherent and consistent manner. N. F. Noy, and D. L. McGuinness in [3] describe the need for ontology as:

- To share common understanding of the structure of information among people or software agents.
- To enable reuse of domain knowledge.
- To make domain assumptions explicit.
- To separate domain knowledge from the operational knowledge.
- To analyze domain knowledge.

At present, the main methods of ontology construction are: TOVE, Skeletal, METHONTOLOGY, KACTUS, SENSUS, IDEF5 and Seven Steps method. Ontology development process is an iterative process that will continue in the entire life cycle of the Ontology. An ontology is typically built in more-or-less the following manner. The basic steps for building Ontology are [3]:

- Determine the domain and scope of the ontology.
- Consider reusing existing ontology.
- Enumerate important terms in the ontology.
- Define the classes and the class hierarchy.
- Define the properties of classes—slots.
- Define the facets of the slots.
- Create instances.

### III. HUMAN EMOTION CONCEPTS AND CONCEPT MEASURES

Various multimedia applications like sports, news, cinema or any video captured at a function/occasion tries to captures

the emotions of the humans. All these applications need to retrieve the video objects based on the human emotions. The proposed approach is generic flexible and not specific to any video application. It is designed and constructed for all videos where human beings and their emotions are been captured. In sports video sports players, sponsor etc would like to see all the video objects where the audiences are happy, overwhelm, sad etc (score, out, amazing shots). In cinema the user would like to watch the video objects with emotions like of comedy, sad, compassion, pathetic, furious, anger, heroic, energy, terror, horror, astonishment, surprise, tranquility etc. In news domain, the news reader would like to display the videos of the news clippings pertaining to the emotions of the human as mentioned. Sports video ontology [4], the concept based video retrieval system for sports video explore the method of constructing concept ontology for sports video by identifying the concepts, concept hierarchy and the relations ships. The concepts like events, actions, players etc are identified and represented in an ontology. The current research on construction of ontologies is focusing on identification of concepts like events, actions, objects, locations, and people. But most of the video retrieval requirements are pertaining to the human and their emotions involved in the video. Semantic video retrieval efficiency can be increased by considering the emotions of the humans involved in the video. The semantic video retrieval system will be more effective, if the retrieval system supports the retrieval of video objects or images based on the human emotions.

#### A. Human Emotions - Aesthetics

The ancient scriptures describe nine fundamental emotions from which all complex emotions may be produced. Just as all shade of colors are produced from basic RGB-three primary colors. In the same way all emotions are said to be derived from principal emotions known as Navarasa (in Sanskrit). Sanskrit, an ancient language of India, is also one of the oldest languages in the world. Sanskrit is a member of the Indo-European language family and is a classical language of India. The word Sanskrit is derived from 'sam' which means 'together' and 'krtam' which means 'created'. 'Sanskrit' together means completed, refined and perfected. Nava means 'Nine' and Rasa signifies 'mood,' 'emotion,' 'expression' or 'sentiment.' The Navarasa - aesthetics in the scriptures refer to the nine expressions that humans often show. The long standing concept of Navarasa is a way to express the emotions of human that is exceptionally original. The individuality of the characters is the element that each character is the personification of one rasa-emotion. Their nature, the intensity of their reactions, their strengths, their failings - all guided by the rasa they represent, which in turn plays an important role in Video semantic retrieval. Video objects which are retrieved based on the Navarasa makes the retrieval system highly efficient and the only one of its kind ever made. Navarasa is accepted worldwide and been used in all art forms. Navarasa are the emotions that human show according to the situations.

The Nine Moods - Aesthetics (Nava Rasa) are:

- Shringar – Love, Attractiveness, Amour

- Hasya – Comic, Laughter, Mirth, Comedy
- Karuna – Sadness, Kind-heartedness or Compassion, Mercy, Pathetic
- Raudra – Furious, Anger, Fury
- Veera – Heroic, Courage, Energy
- Bhayanak – Terrible, Fear, Terror, Horror
- *Bibhatsam* - Vibhats , Disgusting, Odious
- Adbhuta – Wonderment, Amazement, Astonishment, Surprise
- Shanta – Peace, tranquility.

In addition to the nine Rasas, two more appeared later in literature are, Vātsalya - Parental Love and Bhakti - Spiritual Devotion.

NavaRasa plays a significant influence on Indian cinema. The *Rasa* method of performance is one of the fundamental features that differentiate Indian cinema from that of the Western world. In the *Rasa* method, empathetic "emotions are conveyed by the performer and thus felt by the audience," in contrast to the Western Stanislavski method where the actor must become "a living, breathing embodiment of a character" rather than "simply conveying emotion."

All videos which involve human beings implicitly convey the emotions. In order to build a more effective retrieval system, human aesthetics-emotions and the intensity of the emotions should be identified, represented, stored and should be used during the video object retrieval.

The queries from different application domains that involve human emotions like:

Sports: Show all incredible shots of Sachin.

Cinema: Show all comedy clippings of Mr. Bean.

News: Show all terrible scenes of xxx disaster, etc.

### B. Concept Measure

All human emotions inhabit different levels of intensity or extent. Or the concept like beauty can be measured as more beautiful, most beautiful, less beautiful, least beautiful. The intensity of the emotion courage can be measured as less, good, very good, incredibly courage. The intensity of the emotion is defined as Concept Measure. All degree adverbs like good, bad, high, low, very, most, more, less, incredibly, worst, bad, excellent, quickly, loudly, awfully, fairly, quite, really, extremely, pretty, fully, almost, little, too, partial, completely, adequately, immensely, perfectly, strongly etc are concept measures. To describe the human emotions, events and actions more effectively concept measures are attached to the concepts. This can be represented as, humans involved in the video, emotions and the intensity of the emotions. Actions and the intensity. Concept measures like easy, good, bad, worst, high, low etc are purely human judgment. And always there is a chance that two individuals may judge it differently. A particular scene may be too comedy to one individual which is just comedy for other individual. The degree of judgment varies from one individual to other. The fuzziness involved in human judgment pertaining to the concept measures are

represented using the fuzzy logic see figure 1 and 2. Concept measures are ranked on a scale of 0 to 1. The descriptions (d) can be described using concept measure (cm) as:

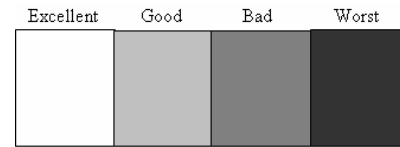


Figure 1. Crisp Sets

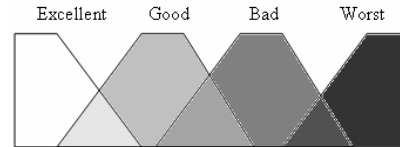


Figure 2. Fuzzy Sets

$d1 = \{(\text{excellent}, 1), (\text{good}, 0.75), (\text{average}, 0.5), (\text{bad}, 0.25), (\text{worst}, 0.0)\}$ .  $d2 = \{(\text{very high}, 1), (\text{high}, 0.8), (\text{medium}, 0.6), (\text{low}, 0.4), (\text{very low}, 0.2), (\text{ground level}, 0.0)\}$ . The queries from different application domains that have human emotion and the intensity of the emotions like:

Sports: Show all excellent shots of Sachin.

Cinema: Show all, incredible comedy clippings Mr. Bean.

News: Show most terrible scenes of xxx disaster, etc.

## IV. A PRACTICAL IMPLEMENTATION

[5][6][7] [8] describes ontology tools. We have used Protégé as an Ontology developing tool to represent the human emotion concepts. Protégé [9] was developed by (<http://protege.stanford.edu>) Mark Musen's group at Stanford University. The human emotions are represented in the Aesthetics ontology using Protégé. Aesthetics ontology and onto graph plug-in representation is shown in Fig.3, Fig.4, and Fig.5. We selected OWL, as the ontology language, which is standard ontology language recommended by W3C.

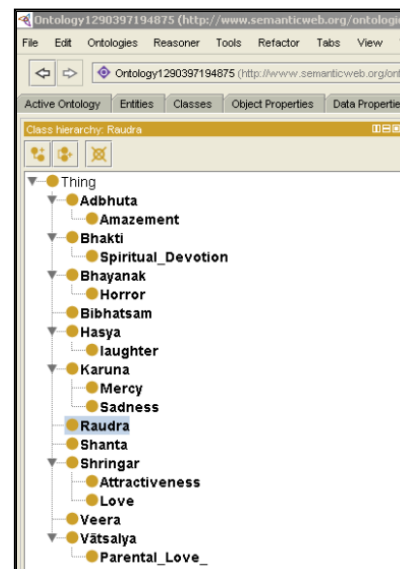


Figure 3. Aesthetics Ontology

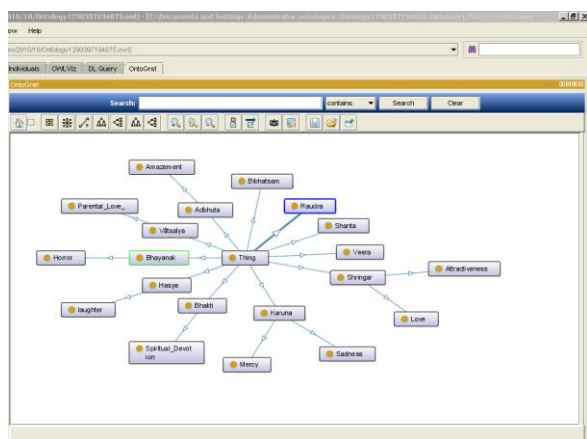


Figure 4. Onto Graph notation of Aesthetics Ontology

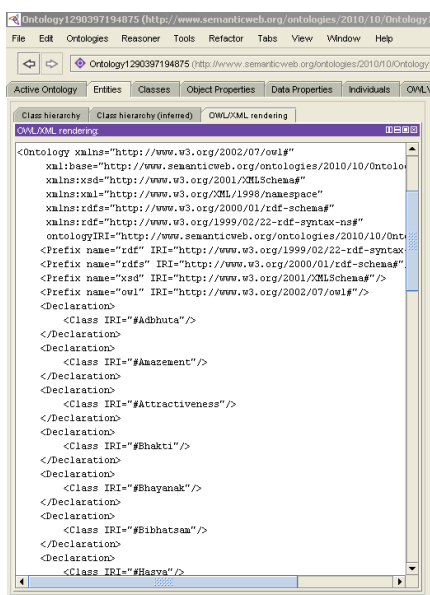


Figure 5. OWL/XML rendering of Aesthetics Ontology

## V. CONCLUSION AND FUTURE WORK

Due to the rapid growth of the voluminous video repositories, the demand for efficient and effective video storage, semantic annotation, indexing and retrieval systems is growing day by day. Domain ontology plays an important role in information retrieval, knowledge representation, knowledge sharing, knowledge integration, knowledge reuse and so on. Ontologies focusing only on the concepts like persons, things, events, actions, places are not complete. Since most of the videos are captured to capture the human and their emotions. The video storage, annotation and the retrieval system would be complete by considering the human emotions and intensity of the human emotions. This raises the need for identification, classification and representation of human emotions. As

Ontology is a kind of concept model that could describe system at the level of semantic knowledge. It aims to access knowledge in a domain in a general way and provides a common understanding for concepts in the domain so as to realize knowledge sharing and reuse among different application programs and organizations. The paper identifies, classifies and represents the human emotions in aesthetics ontology. The intensities of the emotions are drawn using the fuzzy logic. The current paper constructs the aesthetics ontology. Further research could be conducted on semi or automatic identification, extraction, annotation generation of human emotions and the retrieval of video objects using aesthetics ontology.

## ACKNOWLEDGMENT

This work has been partly done in the labs of Adhiyamaan College of Engineering where the author is currently working as a Professor& Director in the department of Master of Computer Applications. The author would like to express her sincere thanks to Adhiyamaan College of Engineering for their support rendered during implementation of this module.

## REFERENCES

- [1] El-Ghalayini, H., (2007). Reverse Engineering Domain Ontologies To Conceptual Data Models, PhD thesis, University of the west of England, UK.
- [2] T.R.Gruber, Toward principles for the design of ontologies used for knowledge sharing, *Human Computer Studies*, 43 (5-6), 1995, 907-928.
- [3] N. F. Noy, and D. L. McGuinness, —Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report, KSL-01-05, 2001.
- [4] Sunitha Abburu, et al, "Concept ontology construction for sports video" Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India 2010, Coimbatore, India September 16 - 17, 2010.
- [5] F. L. Mariano and G. P. Asunción, —Overview and analysis of methodologies for building ontologies, *The Knowledge Engineering Review*, vol. 17(2), pp. 129–156, 2002.
- [6] Cimiano P., Volker J., and Studer R., —Ontologies on Demand? – A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text *Information, Wissenschaft und Praxis*, Vol. 57, No. 6-7. (October 2006), pp. 315-320.
- [7] Duineveld A. et al. —Wonder Tools' A Comparative Study of Ontological Engineering Tools. *Intl. Journal of Himian-Computer Studies*. Vol. 52 No. 6, pp. 11 11-1 133. 2000.
- [8] Michael Denny, Ontology Building, A Survey of Editing Tools, <http://www.xml.com/pub/a/2002/11/06/ontologies.html>.
- [9] Matthew Horridge, Simon Jupp, Georgina Moulton, Alan Rector, Robert Stevens, Chris Wroe. OWL Ontologies using protégé 4 and CO-ODE Tools Edition 1.1. The University of Manchester . October 16,2007.

## AUTHORS PROFILE

Dr. Sunitha Abburi: Working as a Professor and Director, in the Department of Computer Applications, Adiyamaan College of Engineering, Tamilnadu, India. She received BSc and MCA from Osmania University, A.P, India. M.phil and Ph.D from Sri Venkateswara University, A.P, India. She is having 13 years of teaching experience and 3 years of industrial experience.

# Reconfigurable Hardware Intelligent Memory Controller for H.264/AVC Encoders

Kamel Messaoudi

LERICA Laboratory  
Badj Mokhtar University  
Annaba, Algeria

kamel.messaoud@univ-annaba.org

El-Bay Bourennane

LE2I Laboratory  
Burgundy University  
Dijon, France

ebourenn@u-bourgogne.fr

Salah Toumi

LERICA Laboratory  
Badj Mokhtar University  
Annaba, Algeria

salah.toumi@univ-annaba.org

**Abstract**—In this paper, we propose an intelligent memory controller for H.264/AVC CODECs with memory management strategies based on a preloading mechanism in order to reduce the number of accesses to external memory. The controller is used both in simulation and implementation platform for the H.264 encoder. This platform uses an external DDR2 memory to record an image sequence, and an intelligent component with local memories to read the images periodically according to the needs of the H.264 processing modules. Several on-chip memories are used to avoid accessing to off-chip memory and consequently decrease energy consumption. The proposed memory controller can be adapted to the various profiles defined in the standard. The intelligent controller is implemented in VHDL and verified to run at 114 MHz using a Xilinx virtex5-ML501 platform. The proposed architecture occupies 10% of the FPGA's resources and ensures data for the processing modules in H.264 encoder.

**Keywords**—Memory management; intelligent controller; FPGA; H.264/AVC codec; real-time processing.

## I. INTRODUCTION

The new video coding standard recommendation H.264 of ITU-T, also known as international standard 14496-10 or MPEG-4 part 10 Advanced Video Coding 'AVC' of ISO/IEC [1], is the latest coding standard in a sequence of video coding standards: H.261 (1990), MPEG-1 Video (1993), MPEG-2 Video (1994), H.263 (1995, 1997), and MPEG-4 visual or part2 (1998) [2]. The H.264 video coding standard achieves a significant improvement in coding efficiency when compared to other coding methods. It can save as much as 25% to 45% and 50% to 70% of bitrate when compared to MPEG-4 and MPEG-2 respectively [3]. However, the computational complexity of the H.264 encoder is drastically increased, resulting in practical difficulties with its implementation on an embedded platform. In order to use this standard in real-time applications, it is necessary to implement it on hardware accelerators. Moreover, in many cases, it might be necessary to define custom memory management techniques to ensure different data-inputs.

To meet real-time requirements of a large number of multimedia applications requiring video processing, researchers and engineers have presented solutions to implement the codec or to optimize the various processing modules. The memory part that provides the input values to the encoder modules is

not described for all inputs. In fact, most solutions assume that the data-inputs of all modules use a 4 pixels wide bus (e.g. 32bit). However, the recording and management of these pixels is not given or is supposed to be carried out by a software component. In order to define a complete simulation platform, we have proposed in this work a memory management strategy for each processing module of the H.264 encoder. This strategy is implemented as an intelligent memory controller capable of providing the necessary data-input for the H.264 processing modules.

This paper is organized as follows: Section 2 presents an overview of H.264 codecs. We provide a functional analysis of the encoder modules and their mapping on computing memory. Section 3 describes the notion of simulation platform based on a DDR2 memory, especially for the H.264 modules. Section 4 gives an overview of different hardware implementations of the H.264 codec and the proposals for the memory management strategies. Section 5 describes our intelligent controller and the memory management techniques. Section 6 presents experimental results of the hardware implementation using Xilinx platform. The conclusions are stated in section 7.

## II. OVERVIEW OF H.264 CODEC'S

The new visual standard H.264 shares a number of components with previous standards, including the H.263 and MPEG-4. H.264 codec is based on a hybrid model for Adaptive Differential Pulse Code Modulation (ADPCM) and a transformation based on the coding of integers, similar to the discrete cosine transform (DCT) used in earlier standards. This complex coding is performed to take advantage of the temporal and spatial redundancy occurring in successive visual images [4]. Therefore, the H.264 codec combines several efficient algorithms, and the two groups of standardization have paid attention to the concatenation of these algorithms [5]. Indeed, the greatest improvement provided by H.264 is the use of a new intra-coding module to eliminate spatial redundancy in different frames. Another improvement in H.264 is the utilization of macroblocks (MBs) and sub-blocks with different sizes ranging from 4×4 pixels to 16×16 pixels. Moreover, the decoder modules are used in the coding channel in order to manipulate the same frames in the encoder as in the decoder at the cost of increased complexity of processing [2].



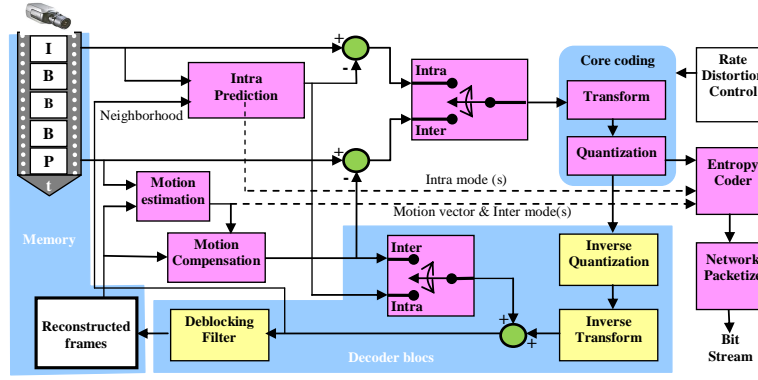


Figure 1. Basic coding architecture for H.264 encoder.

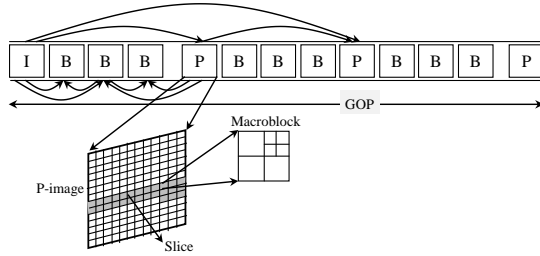


Figure 2. Image organization in H.264 encoders.

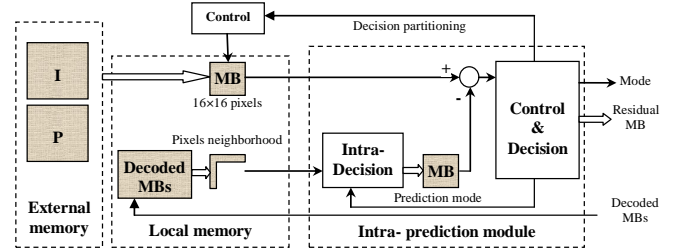


Figure 3. Intra-prediction module architecture and memory management.

#### A. The H.264/AVC Encoding System

The encoding system is composed of the forward (encoding) and the inverse path (decoding). The forward path predicts each MB using intra-prediction or inter-prediction; it also transforms and quantizes (TQ) the residual, then it forwards the result to the entropy encoder module. Finally, it generates the output packets in the Network Abstraction Layer (NAL) module. The inverse path involves the reconstruction of the MB from the previously transformed data by utilizing the inverse transform and quantization (ITQ), the reconstruction module and the deblocking filter [6]. The diagram of the encoder H.264 is shown in Figure 1. In this work, we describe only the modules in direct contact with the external memory.

#### B. Images Organization and H.264 Modules Description

H.264 supports coding and decoding of 4:2:0 progressive or interlaced video. An H.264 video sequence consists of several frames structured as Group Of Pictures (Figure 2). Frames of same GOP are coded according to three methods: intra-coding, inter-coding and bidirectional-coding. A coded frame contains a number of MBs, each containing  $16 \times 16$  luma samples and associated chroma samples ( $8 \times 8$  Cb and  $8 \times 8$  Cr samples). I-macroblocks ( $16 \times 16$  or  $4 \times 4$  pixels) are predicted using intra prediction. P-macroblocks are predicted using inter prediction from reconstructed frames. An inter coded MB can be divided into block partitions of size  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$  or  $8 \times 8$  luma samples, depending on the complexity of the image. If the  $8 \times 8$  partition size is chosen, each  $8 \times 8$  sub-macroblock may be further divided into sub-macroblock partitions of size  $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$  or  $4 \times 4$  for luma samples (and associated chroma samples). Finally, B-macroblocks are predicted using the inter-prediction from reconstructed reference frames ( $I'$  and  $P'$ ) [5].

#### B1. Intra Prediction module

In the intra prediction module, a predicted MB (P-MB) is formed from encoded and decoded pixels (extracted from unfiltered MBs), and subtracted from the current MB. For the luma samples, a P-MB is formed for each  $4 \times 4$  block according to nine optional prediction modes, or for all  $16 \times 16$  MB according to four prediction modes. For the chroma components there are four modes only. The encoder selects the prediction mode for each MB that minimises the difference between P-MB and the MB to be encoded [6]. Figure 3 shows the memory requirements of the intra-prediction module. The control module is specific to the choice of the partitioning decision and the choice for the prediction mode that minimizes residual blocks and MBs. The output of the intra prediction module is a prediction mode for the entropy coding module and the residual MBs for the transformation and quantization modules.

#### B2. Inter Prediction Module

The purpose of inter prediction coding (Motion estimation and compensation) is to predict the current P-frame based on reference frames. The output of this module is a residual MB and motion vectors. The residual MB is coded by the transform and quantization modules, but motion vectors are transmitted directly to the entropy coding module. Figure 4 shows the operating principle of the inter-prediction module and its memory requirements. Motion estimation is widely used to eliminate temporal redundancy between frames in a video sequence. The purpose of motion compensation is to provide additional information to the decoder modules to reduce the residual energy needed to predict the next image. Therefore, the motion compensation module calculates residual MBs [5].



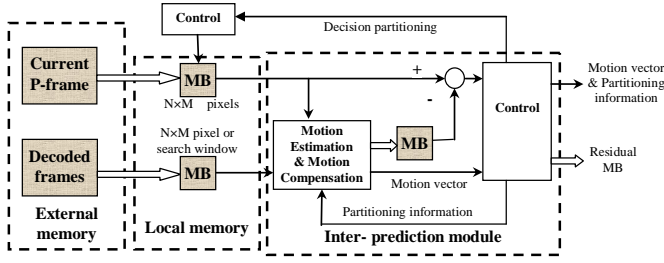


Figure 4. Inter-prediction module architecture and memory management.

In former standards such as MPEG-4 or H.263, only MBs of the size  $16 \times 16$  and  $8 \times 8$  are supported. A displacement vector is estimated and transmitted for each MB using the already encoded and decoded reference frames [3]. The H.264 standards include the support for a range of MB sizes (from  $16 \times 16$  down to  $4 \times 4$  pixels) and fine subsample motion vectors (quarter-sample resolution in the luma component) [7]. These partitions of MB give rise to a large number of possible combinations (Figure 5). A separate motion vector is required for each partition or sub-macroblock [6].

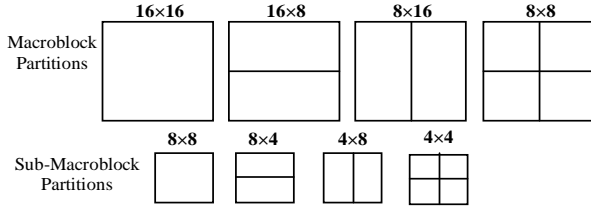


Figure 5. Partitioning of a macroblock and a sub-macroblock.

### B3. The deblocking filter

The deblocking filter module receives as input the reconstructed MB, from the Inverse Transform/Quant (IT/IQ) module. The IT/IQ module generates the reconstructed MB, one  $4 \times 4$  block at a time. The filtering stage is applied to each  $4 \times 4$  luminance and chrominance block boundary within each MB, in a specific order. The deblocking filtering process consists of modifying pixels at the four block edges by an adaptive filtering process. According to the implementing method, the blocks at the output of the filter are not in the same order as the input. It is necessary to add an addressing system to rearrange the processed blocks in their correct positions in the different MBs, and this is done by calculating their corresponding addresses in external memory.

### C. Profile and Levels

H.264 defines a set of three profiles, each supporting a particular set of coding functions (Figure 6). The baseline profile supports intra and inter-coding (using I-slices and P-slices) and entropy coding with context-adaptive variable-length codes (CAVLC). The main profile includes support for interlaced video, inter-coding using B-slices, inter coding using weighted prediction and entropy coding using context-based arithmetic coding (CABAC). The extended profile does not support interlaced video or CABAC but adds modes to enable efficient switching between coded bit streams (SP- and SI-slices) and improved error resilience (Data Partitioning) [2].

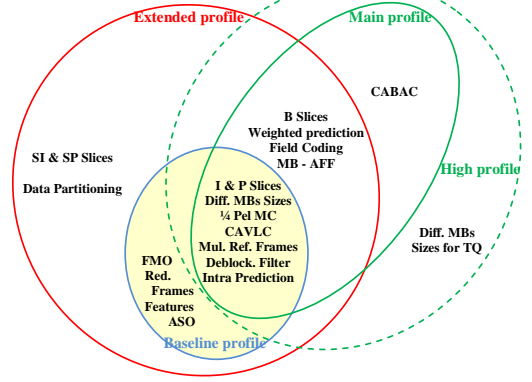


Figure 6. The H.264/AVC profiles and corresponding tools

### D. The Encoder Modules and Memory

Figure 7 presents H.264 modules which make direct use of the external memory and their local memory requirements. In this figure, we note that the different modules need data-input with minimum size of  $4 \times 4$  pixels, which represents a block of pixels of 128bit. In this work we seek to exploit the characteristics of DDR2 memory with an input/output of 128bit and the characteristics of local memories in the ML501 platform that can be configured up to 64/128bit. The input of the deblocking filter, intra and inter modules are therefore considered as 128bit. Even in an MPSoC where the bus size is limited to 32bit, we propose that the controller is directly connected to the processing modules without going through the bus.

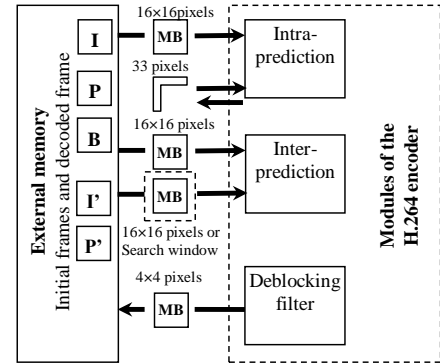


Figure 7. Modules in direct contact with the external memory.

### III. SIMULATION PLATFORM FOR H.264 CODEC

In Figure 9, an implementation platform especially for codecs is presented. The images are initially captured, then stored in memory, processed and finally displayed. It is also possible to store intermediate parameters and intermediate images during processing. Processing stages of the codec are generally applied to MBs of different sizes of recorded frames. To process an image in H.264 codecs, it is necessary to save previous images in the same sequence, which requires simultaneously recording multiple images. Therefore, it is necessary to use an external memory in a prototyping platform with many hardware components to minimize completion time.

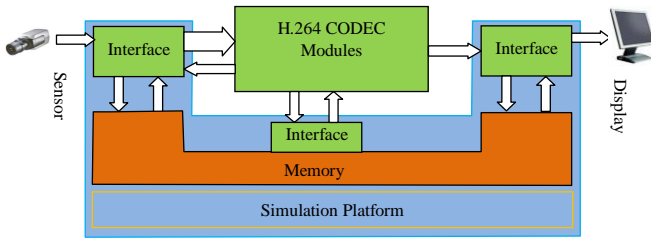


Figure 8. Simulation platform for H.264 modules.

#### A. Prototyping Platform

The proposed intelligent controller is implemented using the Xilinx Virtex5 ML501 development platform [8]. This prototyping platform contains several internal and external components: serial port, VGA output, and other interface resources with the FPGA core type XC5VLX50-1FFG676. It was also fitted with a DDR2 SDRAM SODIMM module of 256MB. The FPGA in the ML501 platform contains multiple IP blocks such as RAMs 36 Kbit blocks configurable according to application needs, FIFO blocks, programmable logic blocks, DSPs, clock generators. This gives the possibility for several configurations of the platform.

The DDR2 memory is organized into 4 banks. Each bank contains a matrix of cells with 210 columns and 213 rows; each cell is composed of 64bit. Therefore, for the luminance signal, each cell may contain 8 pixels. Similarly, a line can contain  $8 \times 210$  pixels, and an image of  $256 \times 256$  pixels occupies 8 lines in DDR2 memory [9]. With this architecture of DDR2 memory, it is not recommended to read an amount of less than 16 pixels.

#### B. Memory Management

Generally, in simulation platforms, the local memory organization is of special importance because it provides temporal storage of data among the codec modules. Usually in video encoder implementations, local memory is used for recording the coefficients and intermediate data, and the external memory is reserved to record image sequences and output files. Image sequences are stored in external DDR2 memory after gathering packets from local memory. Next, the DDR2 memory loads the local memory periodically with blocks and MBs according to the processing module and the image type.

For the intra\_16×16 module, the local memory must contain 16×16 pixels (MB in I-frame) and the 33 neighboring pixels extracted from MBs already encoded and decoded as shown in Figure 7. For the intra\_4×4 module, the memory must contain the 4×4 pixels and the 13 surrounding pixels. In the inter-prediction mode, the memory must contain 16×16 pixel MB in the current P-frame and 16×16 pixel MB in decoded I-frame and P-frame. This last MB must be reloaded periodically from a search window according to the scan mode. There are two appropriate ways to create this architecture:

- In one step, the search window from the external memory is downloaded to the local memory. This has the disadvantage of requiring local memory.

- Use only 16×16 pixel memory, and load this for each inter processing step. This has the disadvantage of requiring a large number of accesses to external memory. To solve this problem, two memories of 16×16 pixels can be used, instead of single-mode ping-pong. The ping-pong memory check doubles the memory capacity.

#### IV. OVERVIEW OF MEMORY CONTROLLER IMPLEMENTATIONS

In the literature, there are many hardware architectures proposed for the H.264 modules. There is a plethora of results related to the design and the implementation of H.264 encoders and of H.264 decoders, but there are only a few published papers that describe the simulation platform and the memory management which provides the input for different modules. Such data-inputs are needed to obtain complete hardware H.264 encoder solutions.

The authors in [10, 11, 12] proposed methods for memory management specifically for the H.264 decoder. Memory management is limited to store the reference image necessary to reconstruct other images in the same sequence. In the decoder, memory access is sequential and limited to reading the MBs one after another. Conversely at the H.264 encoder, the MBs from several frames overlap. In [13] the authors used multiple memories (12 single port memories) to load the module of motion estimation with the image pixels, but the method used to load these memories is not mentioned. In [14] the authors presented an implementation technique to optimize the H.264 encoder on an embedded symmetric multiprocessor architecture. Finally in [6] the authors proposed a complete implementation of the H.264 encoder using a pipelined architecture, and all the modules were optimized with respect to the VLSI cost. In these two papers, authors propose a new technique to optimize memory management to exploit the memory subsystem on an embedded platform, but the supplier and the architecture for the platform and the memory management are not provided.

In this work we propose a complete simulation and implementation platform for the H.264 encoder with an intelligent controller for memory management. This platform will enable us to propose and verify different hardware architectures to implement the H.264 encoder modules.

#### V. THE PROPOSED INTELLIGENT CONTROLLER

Using an external memory to record the sequence of images, requires an intelligent module to read blocks of pixels from it and to feed them to the encoder processing modules. This intelligent module consists of a standard memory controller and an intelligent part customly designed to calculate addresses of MBs necessary for each processing module. Moreover, the controller has to synchronize different accesses to the external memory, and that according to a defined order and in predefined times. In the intelligent part, we must also use any information of the H.264 encoder and any information concerning data organization in each processing module. In Figure 10, we propose hardware architecture for an intelligent memory controller especially designed for H.264 codecs.

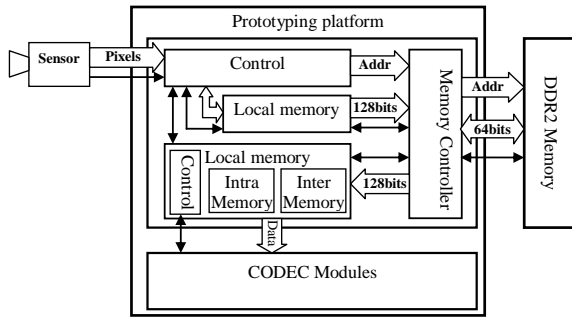


Figure 9. Fig. 10. Proposed memory management for H.264 codecs

Note that the separation of the intelligent part and the initial memory controller is important in order to make our intelligent controller adaptable to several memory types. In fact, to adapt the hardware of a well-defined memory, via the intelligent controller, simply implies to change the standard memory controller. The intelligent part does not change no matter the type of external memory. The separation of the address calculation part and the processing modules is also very important as the module only handles the processing and not the memory address generation. This architecture also allows us to identify and to maintain the size of the input bus for each module of the encoder, and this according to the needs of each processing unit and the chosen parallelism.

#### A. Functional Description of the Controller

Figure 10 shows the use of two types of memory (internal and external). The external memory is used to record the sequence of images, and the internal memory is reserved solely for recording MBs being processed.

- Initially the input pixels (8bit or 32bit) are collected by the control unit in a first internal memory, in packets of 128bit,
- Then these packets are transmitted to the external memory. Address and control signals are provided by the control unit. These first two steps are repeated periodically and continuously with the flow of pixels at the input.
- At the end of each image recording step in the external memory, the control module sends a signal to the processing modules to make a decision on the type of processing to perform.
- The reading portion consists of a local memory to save current MBs, and a reading control part that provides synchronization between the writing and reading of pixels.
- The control part also incorporates an intelligent reading module, to ensure reloading of the internal memory from external memory by MBs, and that following the request of the read control module. The address calculations and the control signals for external memory are provided in this section.
- The processing modules are periodically available in the internal memory and deal only with processing.

These modules need to read the data only from part of the read control module.

#### B. Intra Memory Description

For the Intra\_4x4 module, we suggest reading four lines of the I-frame in the internal memory (4xlines), where each line contains M pixels. Therefore, this memory contains M/4 blocks. In addition to this internal memory, we use another two memories to store neighboring pixels, a memory (1xline) with (M+1) pixels to store upper neighborhood pixels, and a memory (4xpixels) with four pixels to store left neighborhood pixels (Figure 11).

- The 1xline memory is initialized with zeros for the first line of blocks. At the end of the processing of this line, the 1xline memory is completed by the last line of pixels in 4xline memory, to serve as the upper neighborhood pixels for the next line of blocks.
- The 4xpixels memory is initialized with zeros at the beginning of the processing for each line of blocks. At the end of processing, this memory is charged with the four right pixels of the block encoded and decoded, to serve as the left neighborhood pixels for the next block in the same line.

Using two additional memories in this architecture, the number of accesses to the DDR2 memory is reduced. For the intra\_16x16 mode we propose the same principle using 16 lines of pixels, the MB is processed block-by-block in 16 times.

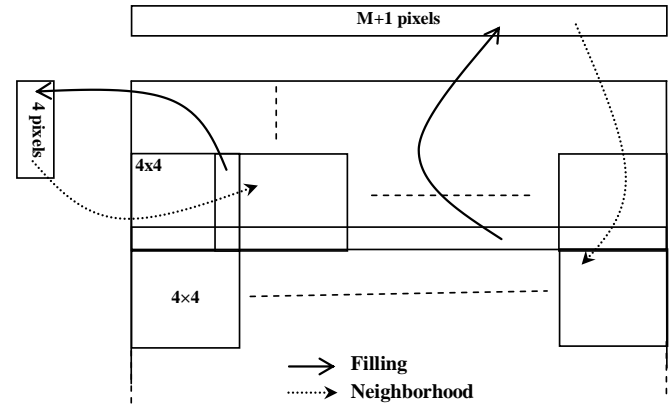


Figure 10. Description of Intra\_4x4 MBs and their neighborhood pixels.

#### C. Inter Memory Description

For the inter prediction module, we suggest reading the central MB from the current P-frame and the search window from encoded and decoded frames. After receiving information from the processing modules (search mode and window size), the controller will determine the type, the number, the size and the order of blocks and MBs to be transmitted.

For hardware implementation, we propose (as with the intra module) to read 16 lines of the current P-frame at once, which gives M/16 MBs, and 16 lines in decoded frames to construct the search windows for each P-MB. Discounting the first reading where 16+d lines must be read, 'd' is the maximum displacement.

#### D. Summary of Rational for the Intelligent Controller

The idea of the intelligent controller for H.264 codecs is to separate the processing modules part from the memory management part, or to design an intelligent module that periodically supplies the processing modules with the required data at a time. Adopting this technique, the H.264 modules manipulate different frames stored in the external memory via a local memory that contains only MBs being processed (Figure 12). In order to perform this separation in processing module and memory management, it is necessary to implement each of the H.264 modules. Moreover, it is important to understand the different read scenarios and the available possibilities for parallelism. These modifications result in a platform for simulation of the codec, in such a way that it is possible to verify the architectures of the processing modules separately.

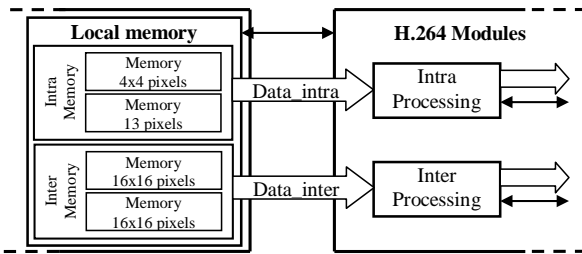


Figure 11. Idea summary

#### E. Architectural Description of the Controller

Figure 13 shows the software architecture chosen for the implementation of the H.264 encoder. The proposed architecture is comprised primarily of a simulation file, which in turn contains two main components:

1. A first part only for simulation composed of two memory modules to record image sequences.
  - A model of SDRAM memory to record the input images.
  - A model of DDR2 memory to save images in output.
2. A second synthesizable part represents the implementation of the H.264 encoder. This part is composed of two main components:
  - The intelligent memory controller which allows images to be recorded in the DDR2, and then loads the MBs into the local memory.
  - The second part represents hardware implementations of different H.264 encoder modules.

The first part with the intelligent controller represents a simulation platform for the H.264 encoder. This platform allows the simulation of various encoder modules. The intelligent controller part consists of several components:

- A DDR2 controller that ensures the communication with the DDR2 memory.
- A (write\_image) module to read pixels in the input and to load the local memory (4x128bit), then transfer them to the DDR2 memory.

- A (read\_blocks) module that calculates address and control signals and that reads the MBs from a local memory
- A (write\_blocks) module to save MBs already processed, decoded and filtered.
- A control module to synchronize various operations and different memory accesses.
- Local memories to record MBs for the processing modules. This memory is the data-input for various encoder modules.

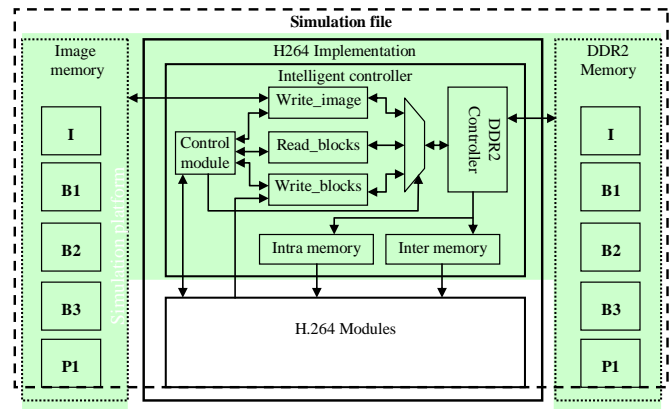


Figure 12. Software architecture for H.264 encoder.

#### F. Different Uses of the Proposed Controller

The proposed intelligent controller can be used directly to feed the H.264 encoder modules with data-input required for each processing. This is to exploit the DDR2 memory features and to give inputs of 128bit to obtain maximum opportunity for data parallelism. Also, the controller can be connected to a data-bus with other components including hardware accelerators (Figure 14). In both cases, the intelligent controller with the local and external memories is considered as local memory that contains the MBs of different frames. In general, to adapt the controller to an image processing application, it suffices to simply define the signal controller links with this pattern. At the controller, nothing needs to be changed except the outputs of the two local memories.

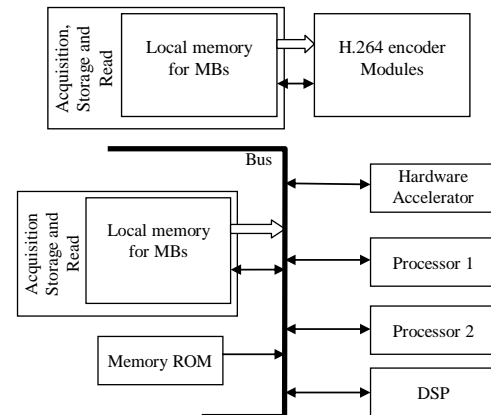


Figure 13. Different uses of the proposed controller.



### G. The Intelligent Controller in an MPSoC

The intelligent controller is the first part in a complete project to incorporate the H.264 encoder in an MPSoC, with a hardware part that contains the modules occupying the largest percentage of the processing time and a software part for the other modules.

In the H.264 encoder, the first three modules in direct connection with the external memory (Figure 1), requiring the maximum size input data (Multi Data input) to provide parallel processing. In a classic design of an MPSoC, these modules are implemented in hardware and are connected directly to the main bus with a maximum capacity of 32bit. This implies that the 128bit data at the output of the memory will be transferred several times, with the following disadvantages:

- The input data for the processing modules is limited by the size of the bus, which limits data parallelism.
- The main features of DDR2 memory (reading bits in packets) are not exploited, which implies that our controller is faster than the processing modules.
- The main bus of MPSoC will be overloaded with data transfer alone.
- The FIFO of memory and the FIFO of the controller will be overloaded by data which hampers the smooth functioning of the controller, which also ensures the operations of recording pixels and the input of already treated MBs.

To overcome these drawbacks, we propose that the controller be directly related to processing modules incorporated into hardware to provide a Multi-Data architecture. The results of these modules can be transferred via a different bus to other hardware modules or via the main bus to the processor. The entire controller and the processing modules will be considered by the processor as one or more hardware accelerators.

## VI. EXPERIMENTAL IMPLEMENTATION RESULT

The proposed hardware architecture for the intelligent memory controller has been described in VHDL language. We used ModelSim for simulation and Xilinx-ISE for synthesis, and we used the embedded development kit (EDK) to integrate the controller in a system based on reconfigurable soft-core processor. For simulation, a test bench file is written to replace the image capture part. The file must provide the image pixels at a frequency (fpixels) calculated according to the type of the input video sequence format CIF with 352×288 pixels/image and 10 images/second.

At the input of the controller, the pixels are collected into groups of 16 pixels in 4 times ( $128\text{bit} \times 4 \Rightarrow$  Burt mode) at frequency fpixels, then the 64 pixels (512bit) are transferred to the DDR2 memory through the memory controller at the FPGA frequency. Figure 15 shows the transfer timing to the DDR2 memory of the pixels of an image.

Reading from external memory to local memory is performed periodically as required by the H.264 modules. The intra\_4×4 module is needed for each intra prediction a 4×4

pixel block and 13 neighborhood pixels. The block is loaded directly from the external memory to the local memory and the neighborhood pixels are reused from previous calculations. At the end of each block processing, it saves the pixels that are required for other blocks and MBs. Figure 16 shows the reading of a 4×4 block and the designation of the neighboring pixels. It is similar for other modes of intra-prediction and inter-prediction. Therefore, we propose an intelligent controller that periodically supplies the H.264 encoder modules with blocks and MBs from the sequence of prerecorded images in the DDR2 of the Virtex5-ML501 platform.

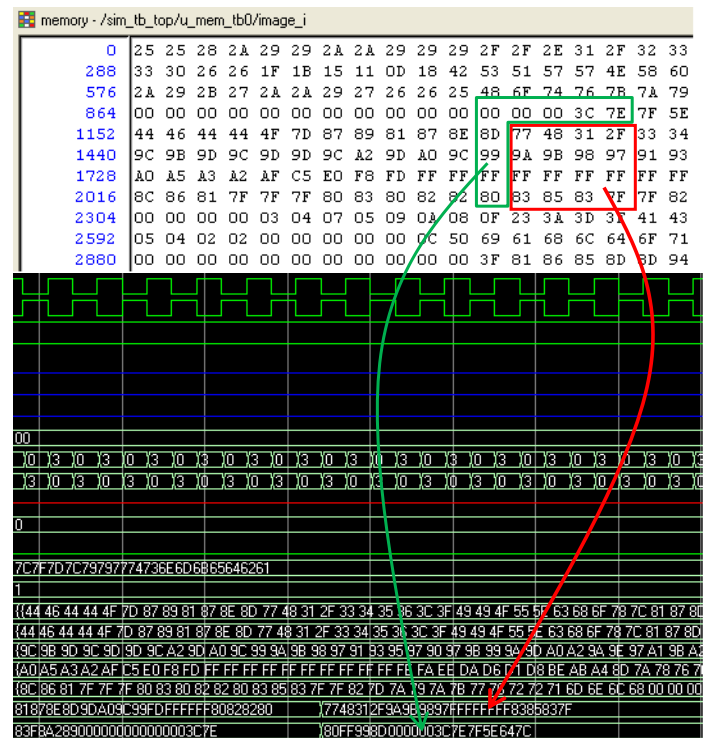


Figure 15. Read memory for 4×4 pixel intra-prediction.

In table 1, we show the local memories used by the H.264 encoder modules:

- The intra-module with two memories Intra\_4×4 and intra\_16×16 (for recording a line of MBs, and a line of upper and left neighborhood pixels), the total memory usage is the memory needed for the two intra-prediction modes and the active memory is the sub-memory transferred to the processing modules for each step.
- The inter module with two memories (the first memory for the current MBs in P-frame and the second memory for the search windows).
- The module of the deblocking filter with a memory to store temporary MBs processed before their transfer to the DDR2 memory.
- A small memory for the collection of input pixels before the transfer to the DDR2 memory in burst mode.

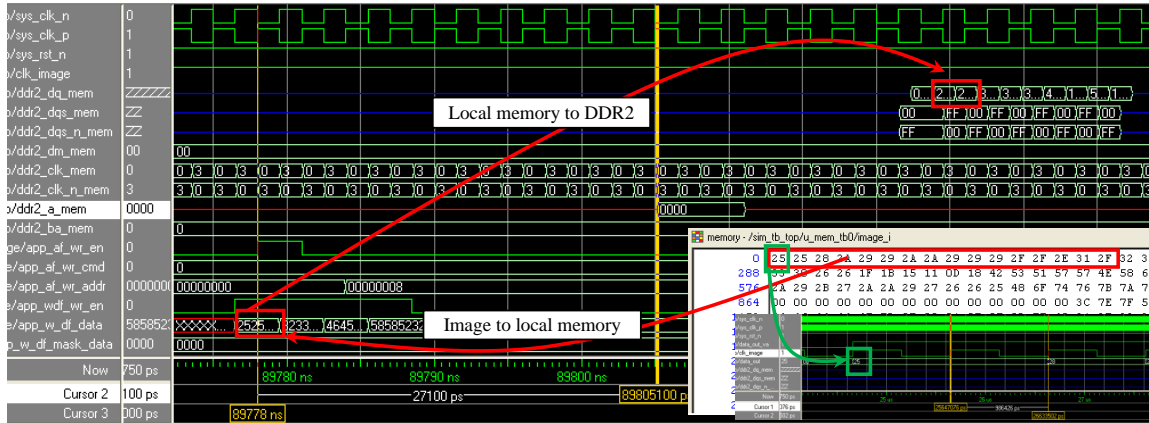


Figure 14. Simulation of the reading of pixels from image to DDR2.

TABLE I. LOCAL MEMORIES USED IN THE INTELLIGENT CONTROLLER

Video format		CIF 352x288	HD TV 1280x720
Used memory			
Read memory		0.500 Kb	0.500 Kb
Intra memory	Intra 4x4	11.281 Kb	28.160 Kb
	Intra 16x16	38.375 Kb	95.750 Kb
	Total memory intra	38.375 Kb	95.750 Kb
	Active memory	(0.226 → 2.257)Kb	
Inter memory	Image P	36 Kb	90 Kb
	Search window	54 Kb	135 Kb
	Total memory inter	90 Kb	225 Kb
	Active memory	(0.25 → 4) Kb	
Deblocking filter memory		9 Kb	9 Kb
Total of used memory		137.875 Kb	330.25 Kb
Percentage of FPGA memory used		7.98 %	19.11 %
Total after optimization		101.875 Kb	240.25 Kb
Percentage after optimization		5.89 %	13.90 %

TABLE II. EXTERNAL AND LOCAL MEMORIES USAGE PERCENTAGE

Profile		DDR2 memory used		Local memory used	
		CIF Image	HDTV Image	CIF Image	HDTV Image
Baseline	Capacity	297 KB	2.637 MB	101.875 Kb	240.25 Kb
	percentage	0.113%	1.03%	5.89 %	13.90 %
Main	Capacity	1.16 MB	10.747 MB	101.875 Kb	240.25 Kb
	percentage	0.453%	4.12%	5.89 %	13.90 %

TABLE III. RESULT OF SYNTHESIS

Device Utilization Summary (estimated values)			
Resources	Avail	Used	Utilization
Number of Slice Registers	28,800	2,860	9 %
Number of Slice LUTs	28,800	3,052	10 %
Local Memory Kb (CIF Image)	1,728	101.875	5.89 %
BRAM/FIFO Block	48	6	12 %

In the same table, the total amount of local memory is given, together with the percentage of memory used in the Virtex5-ML501 platform.

In addition to memories already described, other intermediate local memory must be reserved for other H.264 encoder modules when using pipeline architecture. These

memories are usually allocated to record the remaining MBs with a size of 4x4 pixels.

Table 2 shows the external and the local memory used in the intelligent controller with the percentage of used memory in the Virtex5-ML501 platform. The results are given for two profiles of the H.264 encoder (baseline profile and main



profile) and for two types of image sequences (CIF and HDTV).

Table 3 shows the synthesis results of the intelligent controller on the Virtex5-ML501 platform. The percentage of used area in the FPGA is about 10%, which gives the opportunity to incorporate the encoder modules into the remaining resources of the platform. To map different local memories, six BRAM (12%) are used. The design achieved a maximum clock frequency of 114 MHz on the FPGA board.

## VII. CONCLUSION

In this work, we have proposed an intelligent memory controller for H.264 codecs to build a real-time platform for the simulation and hardware implementation of the encoder modules. A number of optimizations are applied to the controller, to supply the processing modules with data periodically with the required data (blocks and MBs) in each processing step. Simulations and synthesis results were performed with ModelSim and ISE using the Virtex5 ML501 platform. The proposed architecture for the intelligent controller occupies 10% of the FPGA's resources and 12% of the available BRAM, and operates at a maximum frequency of 114 MHz. This study allowed us also to determine the time (the number of clock cycles) required to process each MB according to the manipulated image. The proposed platform with the intelligent controller can be employed in many real-time implementations of the H.264 encoder. It has great flexibility for reconfigurability.

## REFERENCES

- [1] Recommendation ISO/IEC14496-10:2003, "Coding of Audiovisual Objects-Part 10: Advanced Video Coding, also ITU-T Recommendation H.264, Advanced video coding for generic audiovisual services," ISO/IEC, 2003.
- [2] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: Tools, Performance, and Complexity," IEEE Circuits and Systems Magazine, pp.7-28, 2003.
- [3] T. Chen, C. Lian, and L. Chen, "Hardware Architecture Design of an H.264/AVC Video Codec," IEEE Vol.7D, No.3, pp.750-757, 2006.
- [4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions On Circuits And Systems For Video Technology Vol.13, No.7, pp.560-576, 2003.
- [5] K. Messaoudi, S. Toumi, and E. Bourennane, "Proposal and study for an architecture hardware/software for the implementation of the standard H264," CISA'08, in: American Institute of Physics (AIP), Vol. 1019, No.1, pp.536-540, 2008.

- [6] K. Babionitakis, G. Doumenis, G. Georgakarakos, G. Lentarlis, K. Nakos, D. Reisis, I. Sifnaios, and N. Vlassopoulos, "A real-time H.264/AVC VLSI encoder architecture," Springer Real-Time Image Proc, pp.43-59, 2008.
- [7] J. Chong, N. Satish, B. Catanzaro, K. Ravindran, and K. Keutzer, "Efficient parallelization of h.264 decoding with macro block level scheduling," IEEE-ICME, pp.1874-1877, 2007.
- [8] Xilinx "Xilinx Memory Interface Generator (MIG) User Guide, DDR SDRAM, DDR2 SRAM, DDR2 SDRAM, QDR2 SRAM and RDRAM II Interfaces," Document:UG086 (v2.0), 2007.
- [9] Micron Technology Inc. "128MB, 256MB, 512MB: (x64, SR) 200-Pin DDR2 SDRAM SODIMM Features," Document: 09005aef80eec96e, 2005.
- [10] H. Loukil, A. B. Atitallah, F. Ghazzi, F., M. A. B. Ayed, and N. Masmoudi, "A Pipelined FSBM Hardware Architecture for HTDV-H.26x," Springer-International Journal of Electrical Systems Science and Engineering pp.128-135, 2009.
- [11] C. H. Chang, M. H. Chang, and W. Hwang, "A flexible two-layer external memory management for H.264/AVC decoder," IEEE-SoC conference, pp.219-222, 2008.
- [12] S. H. Ji, J. W. Park, and S. D. Kim, "Optimization of Memory Management for H.264/AVC Decoder," IEEE-ICACT, pp.65-68, 2006.
- [13] H. Chung and A. Ortega, "Efficient memory management control for H.264," IEEE, ICIP pp.35-50, 2004.
- [14] K. Ouyang, Q. Ouyang, and Z. Zhou, "Optimization and Implementation of H.264 Encoder on Symmetric Multi-processor Platform," IEEE - Computer Science and Information Engineering, pp.265-269, 2009.



**Kamel Messaoudi:** received the engineering degree in automatic from the University of Annaba, Algeria, in 1997, and the master degree in Industrial Computing and image processing from the university of Guelma, Algeria, in 2000. He is currently pursuing the PhD in electronic and informatics at the University of Annaba, Algeria, and the Burgundy university of Dijon, France. He is Assistant Professor in the department of Computer science, University Mentouri-Constantine, Algeria. His research interests includes: Image processing, embedded system, FPGA design and Real time implementation. He can be reached at kamel.messaoud@univ-annaba.org.



**Bourennane El-Bay** is Professor of Electronics at the laboratory LE2I, (Laboratory of Electronics, Computer Science and Image). University of Bourgogne, Dijon, France. His research interests includes: Dynamic reconfigurable system, Image processing, embedded system, FPGA design and Real time implementation. He can be reached at ebourenn@u-bourgogne.fr.



**Toumi Salah** is Professor of Electronics and Head of LERICA laboratory in Badji Mokhtar University, Annaba, Algeria. His research interests includes: Telecommunications, embedded system, micro-electronic and laser. He can be reached at salah.toumi@univ-annaba.org.

# FRAGMENTATION INVESTIGATION AND EVALUATION IN DISTRIBUTED DBMS USING JDBC AND OGSA-DAI

Ahmed Almadi, Ahmed Manasrah, Omer Abouabdalla, Homam El-Taj

National Advanced IPv6 Center of Excellence (NAv6)

UNIVERSITI SAINS MALAYSIA

Penang, Malaysia

{almadi, ahmad, omar, homam}@nav6.org

**Abstract**— This research investigates and evaluate the impact of the fragmentation on different database retrieval modes based on derived horizontal fragmentation by generating and distributing the query to the servers (distributed search) or send the query to the direct server (direct search). Moreover, it provides recommendation on suitable query execution strategies based on a proposed fitness fragmentation formula. Furthermore, examine the suitable technology such as OGSA-DAI and JDBC in grid database to examine the time overhead in distributed systems and grid environments in different cases like size or number of servers. The results show that the fragmentation's time performance impact is clearly effective and positively applied while increasing the database size or the number of servers. On the other hand, the OGSA-DAI kept on showing slower execution time on all conducted scenarios, and the differences between the execution time exceeds up to 70% while increasing the size of data or number of servers. In addition, this thesis has tested the impact of fragmentation search against the distributed search where the first one submit the query to direct server(s) (direct search), and the second one distribute the query to the servers (distributed search). The result shows that the speed effectiveness of direct search technique in JDBC case is around 70% faster than the distributed search and around 50% faster in OGSA-DAI case.

**Keywords**—component; JDBC; OGSA-DAI; Fragmentation; Distributed DBMS

## I. INTRODUCTION

In distributed systems, data are fragmented, located and being retrieved in a transparent manner among the distributed sites[3]. Therefore, accessing some distributed data from different locations are applied using a “View” of the data. However, technically, in the distributed systems, the catalogue database is an essential demand where it makes an affix for the physical location into the catalogue [11].

Moreover, web-services are playing a big role on retrieving the fragmented database and applying certain services. Fragmentation is considered to be one of the most important phases that is been conducted to achieve the distributed database design. Yet, the impact of the fragmentation performance on the case of increasing or decreasing the overhead is unclear.

Moreover, some noticeable overheads are come into view clearly from several new technologies in the distributed systems such as OGSA-DAI middleware. A reason is that the high-level technologies and processing gives a noticeable overhead. In particular, the perceptible overheads are appearing clearly on retrieving databases and accessing the distributed systems. From this point of view, main research question is "How to reduce the overhead in the grid systems and distributed systems on distributed database retrieval service?"

Sub-questions arise from the research are as the following:

- What is the best database size to apply the fragmentation if we consider the performance in order to generate sub-queries to the servers or just to the local server?
- What is the tradeoff between transparency and the performance in case of using JDBC and OGSA-DAI?

This paper focuses on the impact of the fragmentation on different cases of database systems, and on the JDBC performance under several layers of executions against the OGSA-DAI. The evaluation part will be based a quantitative evaluation and the execution time overhead is the main attribute of the evaluation.

## II. LITERATURE REVIEW

### 1) Distributed Query Performance

In processing such an index partitioning scheme two approaches are presented in response to a range query. A comparison between such approaches and other similar schemes is done in order to compare their performances. Accordingly, this performance is assessed from the perspective of the response time, system throughput network utilization and disk utilization. Taking in account varying the number of nodes and query mix [12].

Sidell in (1997) presented the distributed query processing problem in Mariposa. A comparison of its performance with a traditional cost-based distributed query optimizer is obtained [9].

The ability to adapt a dynamic workload is displayed through a Mariposa system. In addition, the adaptive distributed query processing in Mariposa and its interaction with multi-user workloads, network latencies and query size benefits are investigated. Performance results observed to show that the Mariposa system outperforms a static optimizer as it distributes works regularly among the available sites. Besides, it is to be noticed that the overhead which is introduced by Mariposa's budding protocol gives insignificant results if it is used with large, expensive queries. In addition, though for small queries it is outweighed by the benefits of load balancing. Truthfully, the comparisons based on the TPC-D benchmark show that the authors' point of view in which their approach behave as a static optimizer is influenced by network latency and query size [9].

In 2003, a paper in title of "Distributed Query Processing on the Grid" [10] argues on the significant of the distributed query processing in the Grid and on the facilities in the grid that support the distributed query processing producers. The paper describes a Polar prototype implementation of distributed query processing running over Globus. They used a bioinformatics case study to illustrate the benefits of the approach [10].

Oliveira et al., have presents in 2007 a paper that shows the development on the grid computing and a comparison was conducted on two algorithms for planning the distribution and parallelization of database equerry on grid computing [8]. Showing the partial order planning algorithm with resource and monitoring constraints is the best choice for distribution and parallel DBMS queries was their main contribution.

## 2) Investigating the OGSA-DAI

In grid computing, many investigations and studies on OGSA technology where aim to decipher the importance of OGSA-DAI and the benefits of its services.

An overview of the design and implementation of the core components of the OGSA-DAI project was presented in High-level manner. The paper describes the design decisions made the project's interaction with the Data Access and Integration Working Group of the Global Grid Forum and provides an overview of implementation characteristics. Implementation details could be seen from the project web site [2].

In describing experiences of the OGSA-DAI team, a team has an experience in designing and building a database access layer using the OGSi and the emerging DAIS GGF recommendations [7].

They designed this middleware to enable other UK e-Science projects which need database access. It also provides basic primitives for higher-level services such as Distributed Query Processing. In addition, OGSA-DAI intends to produce one of the required reference implementations of the DAIS specification once this becomes a proposed recommendation and, until then, scope out their ideas, provide feedback as well as directly contributing to the GGF working group [7].

In this paper, issues that have arisen in tracking the DAIS and OGSi specifications are presented. These issues appeared during a development of a software distribution using the Grid

services model; trying to serve the needs of the various target communities; and using the Globus Toolkit OGSi core distribution [7].

In 2008, Hoarau & Tixeuil, presented an experimental study of studying the OGSA-DAI [5]. Results were quite stable and performed quite well in scalability tests, and were executed on Grid5000. It is also discussed that the OGSA-DAI WSI uses a SOAP container (Apache Axis1.2.1) which suffers from severe memory leaks. It is shown that the default configuration of OGSA-DAI is not affected by that problem; however, a small change in the configuration of a Web-service could lead to very unreliable execution of OGSA-DAI [5].

An OGSA-DQP is an open source service-based distributed query processor. The evaluation of queries is supported by this processor. The OGSA-DQP effects over several layers of service-oriented infrastructure. Experiences in investigating the impact of infrastructure layers were discussed in a study in [1]. In addition, this study presents an understanding of the performance issues, identify bottlenecks, and improve response times of queries. It also describes the experiments carried out and presents the results gained [1].

However, as illustrated in Figure 1 the processes in the OGSA-DAI which are in high-level schematically representations are passing through several layers of interfaces between each layer. Therefore, it gives the fact of the time overhead performance through using the OGSA-DAI high-level schematically representation to communicate and retrieve the database [6].

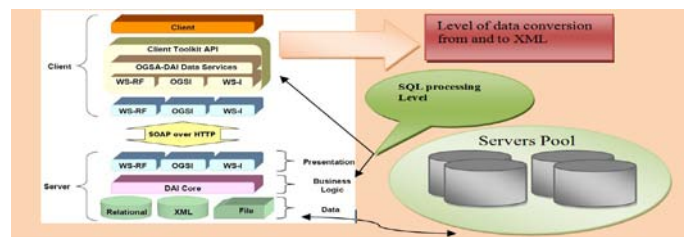


Figure 1. OGSA-DAI architecture and flow processes

## III. FRAGMENTATION FRAMEWORK

The derived fragmentation is a fragmentation where databases are fragmented according to a specific attribute. Since that, a catalogue table is a compulsion in the main server to keep on mind knowledge for all fragmented databases.

Catalogue database is a database that contains the information of the distributed database. It contains the site, the name of the database and the attributes in where the database was fragmented.

We will apply a basic student database which consists of student's information such as name, id and year of each student. Table I. shows the catalogue table which will be conducted in the research implantation.

TABLE I. CATALOGUE DATABASE

Table Name	FABase	Dbname	Serverid	FAconstraints
Student	Year	DB1	S1	1
student	year	DB2	S2	2

Table I. above consists of 5 main attributes that displays the main information about the fragmented relational or tables in the distributed system:

1. Table Name: contains the name of the fragmented tables.
2. FABase: contains the attribute where the tables were fragmented according on.
3. Dbname: contains the name of the DB in the distributed servers where it handles the fragmented table.
4. Serverid: contains the server ID to refer to the server's IP for the distributed DB.
5. FAconstraints: contains the value of the fragmentation base attribute for each table.

#### IV. MAIN FRAMEWORK FOR ENGINE SEARCHING

In distributed systems, database retrieval process is achieved on two main ways. The first way is to directly access the server that contains the required data. The second technique is to distribute the searching query into the distributed servers. In this research, we will be calling these two techniques direct search and distributed search respectively.

The system will take the decision of choosing the searching method after analyzing and understanding the SQL statement. The decision will be based on the existing of the fragmentation attribute via the query, in such case, system will choose the direct search and. The existing of the fragmentation attribute in the query means that the system can get directly the data from the distributed server(s) by getting the site of that server from the catalogue database, by referring to the **FABase** in that table. In this case, the performance will be higher since it reduces lots of time if it was using the distributed search.

From Figure 2. we can see the main architecture of the research framework. The processes can be categorized into five main layers: *User interface*, *SQL Processing System*, *Management Connections (JDBC)* and *Servers pool*.

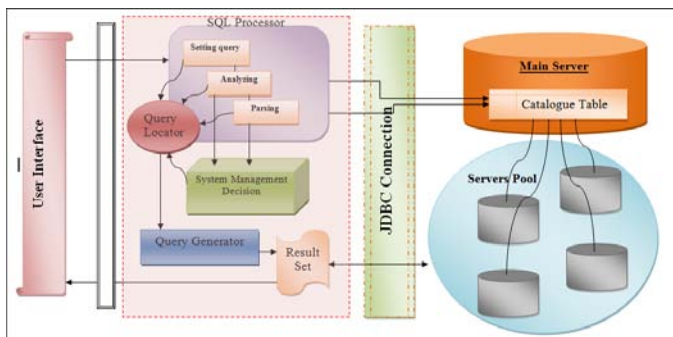


Figure 2. Main framework and architecture

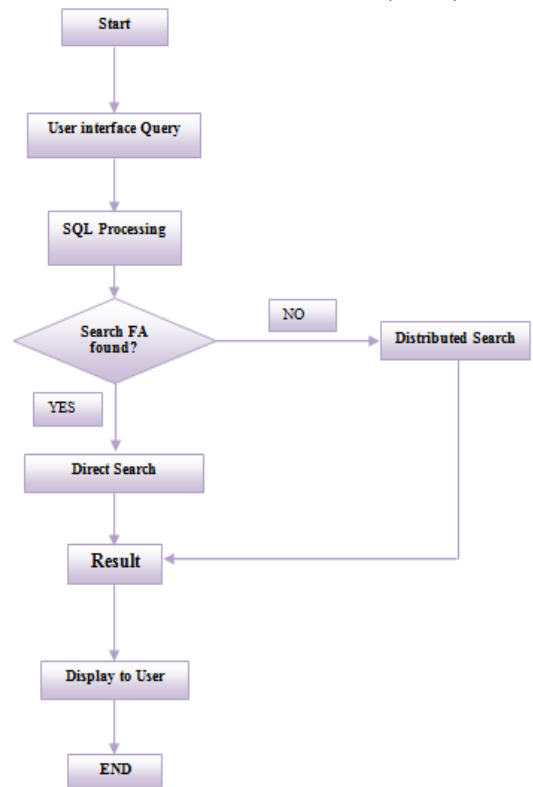


Figure 3. Flowchart for database retrieving based on catalogue database

#### A. SQL Processing Search

In SQL processing step, several procedures will be done inside the engine to process the SQL, which are:

1. *Getting the SQL statement from the web-server*: in the interface, the user will write the SQL statement for retrieving certain data as a string.
2. *Select checking*: (validation) does it start with "SELECT", by using searching method for this word (select), if yes, continue to next step, else, give an error message and back to the main page .
3. *Table Name Searching*: After getting the SQL, a table searching method will be called. Its main job is to search for the table's name inside the SQL statement (string).
4. *Checking Name*: checks if the name was mentioned (the SQL statement is correct so far), if yes, save the table name and continue to next step, else, give an error message and back to the main page.
5. *Retrieving Fragmentation Attribute (FA)*: Get the FA style for that table from the catalogue database which is saved in the main server.
6. *FA Searching*: Another searching method will be called to search for the FA style in the SQL statement.
7. *Found*: If the FA was found in the SQL statement, the system will choose the *direct search*; else it will choose the *distributed search*.



### B. Direct Search

In direct search step, after receiving the FA, several procedures will be done inside the engine to process and achieve the result, which are:

1. *FA Receiving*: System will receive the FA from SQL processing step.
2. *Select Statement Generator*: Calling a generator method that generates a select statement that will select from catalogue database locations according to the FA.
3. *Main Server Connection*: Calling the connection to the main server.
4. *Sending Generated Select Statement*: Send the generated select statement to the main server.
5. *Server Location Retrieving*: Getting back the site information for the server.
6. *Server Connection*: Prepare a connection to that server (site).
7. *Sending User's SQL*: Send the user's SQL statement to the server.
8. *Check Data Availability*: If the data were found, get back the result from the server and save it, else, return a message to the system ("No Data Found").
9. *Getting Results*: Get the result and save to display it to the user.

### C. Distributed Search

1. *Select Statement Generator*: Calling a generator method that generates a select statement, it will select all sites that the table was fragmented to.
2. *Main Server Connection*: Calling the connection to the main server.
3. *Sending Generated Select Statement*: Send the generated select statement to the main server.
4. *Servers Location Retrieving*: Getting back the sites information for the servers.
5. *Servers Connection*: Prepare a connection to all servers (sites).
6. *Sending User's SQL*: Start to send the user's SQL statement to the selected servers at the same time.
7. *Check Data Availability*: If the data were found, stop searching, else, return a message to the system ("No Data Found").
8. *Saving Result*: get back the result from the server and save it.
9. *Display Result*: Display the result to the user.

## V. IMPLEMENTATION AND RESULTS

In the following sections, we will present the process flow or model that the research implementation will be based on.

The main concern on these experiments is to evaluate the usability of JDBC connection in low-level manner against the OGSA-DAI middleware by measuring the time overhead

difference between using distributed processing using JDBC or OGSA-DAI as a quantitative evaluation based. Moreover, examining the significant of the direct search and distributed search approaches will be highlighted.

### A. Implementation

The implementation and examinations were carried out on the **Grid Computing Lab, in University Sains Malaysia, 2008**. Implementation was applied using java code applied on an open development platform (Eclipse). From the hardware point of view, There were 4 workstations were applied to conduct the testing of this research and one for the development detailed in Table II.

TABLE II. HARDWARE USED DURING IMPLEMENTATION AND TESTING

Machine used for development			
Processor(s)	Memory	Hard disk drive	Operating System
Intel® Core™ 2 CPU T7200 @ 2.00GHz	1GB RAM	80 GB	Mac OS X Leopard v10.5
4 Machines used for testing			
Processor(s)	Memory	Hard disk drive	Operating System
Intel Pentium 4 processor 1.8 GHz	512 MB	80 GB	CentOS Rocks

From the software point of view, Software, libraries and web servers where used are:

- o Java EE open development platform (Eclipse).
- o MySQL as the DBMS for the data base.
- o Apache Tomcat 5.5.25 web server. downloaded from: (<http://tomcat.apache.org/>).
- o Main libraries: java.sql.connection; java.util.\*;

In addition, the OGSA-DAI were deployed in the stations.

### B. Experimental Designe

In order to achieve the objectives of the research, several experiments and scenarios has been examined and analyzed. The first group of tests was concerned about comparing the time execution performance of JDBC beside OGSA, besides the speed up while increasing the database size. The second group of tests was concerned on the symptomatic of the of fragmentation service system. Moreover, analysis of the experimental results and recommendation will be presented.

### C. Performance of JDBC next to OGSA-DAI

According to the research framework, several experiments have been conducted based on two general scenarios of distributed database retrieval systems. These main scenarios are generally under the following headlines:

1. Direct Search.

## 2. Distributed Search.

### 1) Scenario 1: Direct Search

In this case, the system will access the data from direct server according to the fragmentation constraint. We based our scenarios on 2 cases:

1. Data quantity in the servers.
2. Server Quantity.

The testing was applied using different connection approaches:

- Request is sent via JDBC in low level of parsing and connecting.
- Request is sent via OGSA-DAI Grid Data Service (GDS) to the same relational database.

#### (a) Case of Data Quantity

Each query case is being tested while increasing the amount of database rows, and the time for each case is given in Milliseconds (ms).

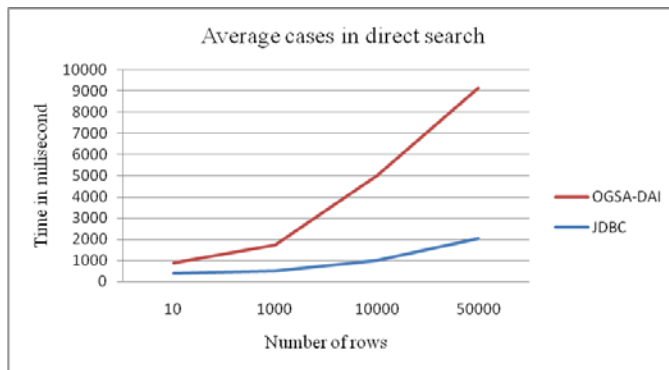


Figure 4. Average query execution time in JDBC and OGSA-DAI

The previous experiments show a less time overhead on using JDBC low level connection rather than OGSA-DAI connection on scalable database size.

#### (b) Case of Server Quantity

In this case, the experiments are done while increasing the number of servers and taking the previous average execution time experiment result..

These results were gained by applying different size of database starting from 10 rows up to 50000 rows. The average execution time of the query on the multiple amounts of servers via JDBC and OGSA-DAI is illustrated in Figure 5.

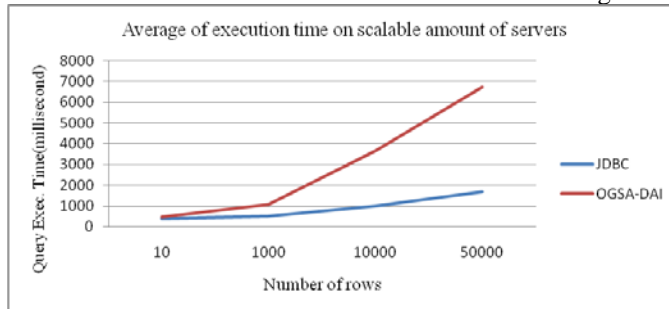


Figure 5. Average query execution time on multiple amounts of servers in direct search

As a conclusion for the previous experiments, results show a less time overhead using JDBC connection in comparison to OGSA-DAI.

### 2) Scenario 2: Distributed Search

In this case, the system will send generated query to the servers at once, and access the data from all fragmented databases. The experiments aims to show the distributed search approach among JDBC and OGSA-DAI. We based the scenarios on the cases of **data quantity in the servers**.

The testing was applied using different connection approaches:

- Request is sent via JDBC in low level of parsing and connecting.
- Request is sent via OGSA-DAI Grid Data Service (GDS) to the same relational database.

Accordingly, we will start to conduct the phase of data quantity in servers, by increasing the data gradually and test the given scenarios of query. From here the computational time for the database retrieving must show differences between JDBC connection and OGSA-DAI.

Figure 6 illustrates clearly the differences query execution time using JDBC and OGSA-DAI, which it gives better results using the JDBC.

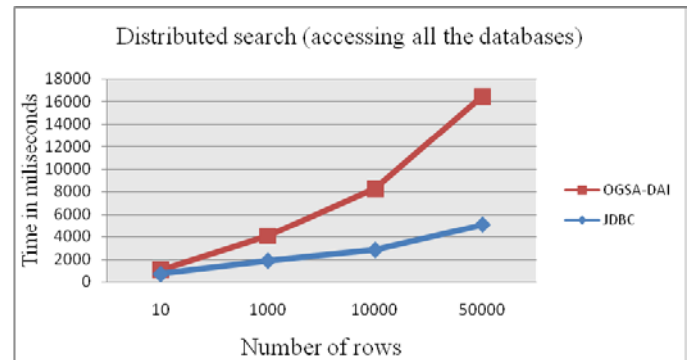


Figure 6. Execution time on case of retrieving the entire database from all servers

The second distributed query is being processed among the distributed system. The table below shows the experimental result of that case.

Second query test: "Select \* from Table where gender='M'".

$$\text{Speed up} = \frac{\text{OGSA-DAI}}{\text{JDBC}} \quad (1)$$

In Figure 6, the result is similar comparing with the previous scenario with slightly small differences. According to the previous gained results, increasing of speed up in JDBC while increasing the size of database is clear as shown in the following figure. The speed up was measured by using the Formula (1) of speed up.

Figure 7 shows the speed up and the increasing of the execution time differences between JDBC and OGSA-DAI among the servers.



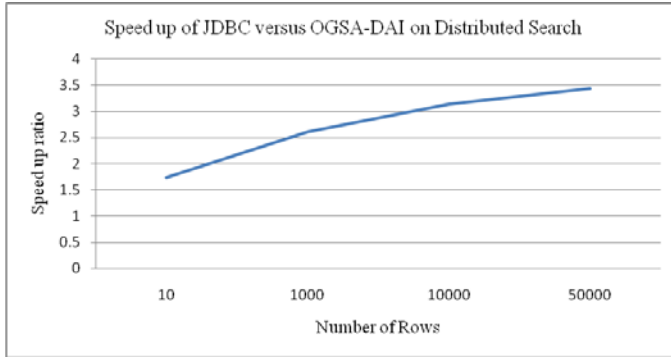


Figure 7. Speed up of distributed process using JDBC versus OGSA-DAI

#### D. Fragmentation Significant

In this case, the test is going to test the execution on multiple servers in order to realize the significant of the fragmentation. In fact, the main goal of these experiments is to answer the question: “*when should we fragment?*”

To answer the earlier question, we need to consider several formulas to be based on while analyzing the answer.

If we consider the following variables:

- Number of servers =  $ns$
- Time of SQL processing =  $tsp$
- Time for retrieving one row =  $tdr$
- Amount of data =  $ad$
- Direct Search Time =  $dst$
- Distributed Search execution time =  $distst$
- Parallel Process time =  $ppt$

The total time  $T$  for accessing data based on distributed database according to general approach is represented by the following formula:

$$T = tsp + \frac{tdr \cdot ad}{ns} \quad (2)$$

From here, time elapsed without distributing the data (local search) gives less overhead than if it was subjected to the distributed system. Accordingly, the execution time for the distributed search is presented as in the following formula:

$$distst = dst + ppt + \text{Join Process time} \quad (3)$$

Consequently, the fitness case here will event when the time for local search is equal to the distributed search time execution for the same amount of data.

From the second formula,  $dst$  search gives 50% less execution time as an average case than the total of  $distst$  average. Base on that, for understanding and analyzing the situations, we presented the following results table that illustrates Formula (3).

By referring to Connolly’s assumptions, that the  $tsp = 70$  milliseconds and  $tdr = 0.005$  milliseconds / one row [4]:

TABLE III. FRAGMENTATION SPEED BASED ON FORMULA (3)

Data amount	1 server (local)	2 servers	3 servers	4 servers
1000	75	142.5	141.6	141.25
5000	95	165	148.3	146.25
10000	120	190	156.6	152.5
30000	270	215	190	177.5

Figure 8 illustrates the result table (Table III) of the fragmentation among the distributed servers. Well, it’s clear that if performance is taking on account, the fragmentation is not always the right choice. Yet, it gives better speed on the case of increasing the database amount and server amount.

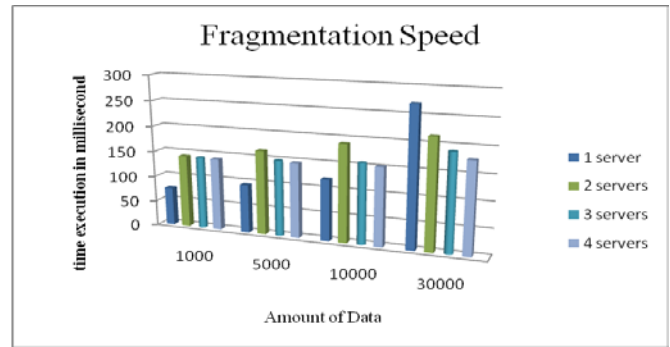


Figure 8. Fragmentation speed on different amount of servers

According to the fitness case, Figure 9 below shows the scale of fragmentation speed on different amount of servers.

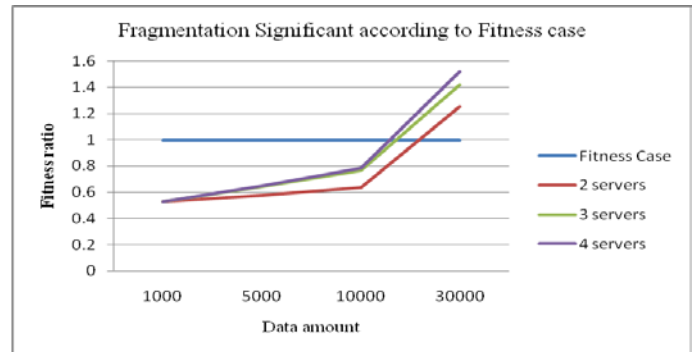


Figure 9. Fragmentation significant based on fitness

From Figure 9 above, it’s clear that the fragmentation scale starts with giving overheads for the system; yet, it shows a better performance on the time when increasing the data amount and the server amount as well.

Figure 9 and analysis table shows quantitative evaluation regarding the proposed fragmentation which uses this amount of database is not giving much efficient results always. The reason is although the generating method is done on parallel and the query sending to the global servers is done on parallel by creating threads, the process spends longer time while generating the obtained methods and creating threads than just

retrieving it from one place. Nevertheless, the results also demonstrate a better result in the case of huge amount of data and comparing more than two servers with each other. The expected approach to gain a good fragmentation result is to exceed the amount and the size of data, and expanding the amount of servers to more than 10 servers will show a better result.

#### E. Direct Search versus Distributed Search

The conducted search schema of direct search is giving a good result in big amount of databases. From previous tests, we can see the average effectiveness that has been found by using the direct search. Figure 4.13 captures the average of these differences while deploying it using JDBC and using OGSA-DAI.

Figure 10 below shows the speed effectiveness of direct search technique. In JDBC case, it shows around 70% faster than the distributed search and around 50% faster in OGSA-DAI case.

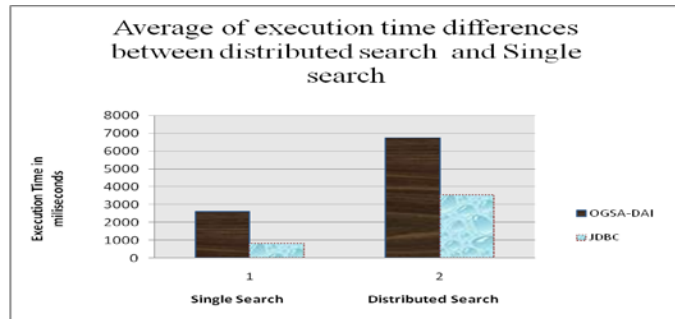


Figure 10. Effectiveness of using direct search versus distributed search

#### F. 4.7 Findings and Recommendations

From the previous presented experiments, we found that:

1. The time overhead performance differences increases between distributed systems processing using JDBC and OGSA-DAI while increasing the data quantity or the number of servers.
2. Nevertheless, the OGSA-DAI still more secure in the grid environment since it provides message and transparent level.
3. Moreover, unlike the distributed processing using JDBC, OGSA-DAI is data heterogeneous.
4. On the fragmentation side, the distributed system was tested on a 4 servers in where the fragmented database where applied there. It will be useful on the event of using a large size of database around more than 10000 in our tests case.
5. Moreover, increasing the database size, gives the fragmentation process a positive impact on the performance.

## VI. CONCLUSIONS

In this research, we have discussed about the distributed systems and the features of the distributed database design. Fragmentation and transparency are a main concept of the

distributed system in high-level manner. Moreover, the approaches, which were conducted to examine the distributed environment, illustrate the distributed concept among fragmentation.

The main framework has passed through different phases to examine the distributed system environment:

1. Fragmentation design.
2. SQL optimizing.
3. In addition, the web-server based approaches to demonstrate the work.

Fragmentation design was based on derived horizontal fragmentation design in this research, and the experimental results on the retrieval system shows a speed up on the distributed search approach using parallel features instead of using central located database. Moreover, the second approach of the direct search shows an efficient execution and less time consuming for retrieval system than distributing the queries among the database.

In the SQL optimizing, the query system shows a good result on processing the query on basic tested cases, where it generates directly a query for sending the request according to the fragmented database.

Accordingly, there is a big relation of distributed fragmentation design and its impact on the query optimization for the reason that the SQL processing system will find a suitable generation method according to the specification of the fragmented tables.

In the web-server, analytical analysis shows that the process under a low-level manner gives a faster speed on distributed systems than the high-level based connection.

As we can see from the results in the previous chapter, the OGSA-DAI gives considerable overheads over the JDBC connection in low level. Moreover, the better results are being gained while increasing the size of database or the number of servers.

## VII. REFERENCES

- [1] Alpdemir1, M. N. Mukherjee, A., Paton1, N. W., Fernandes, A. A. A., Watson, P., Glover3, K., Greenhalgh, C., Oinn, T., and Tipney, H. (2005) Contextualised Workflow Execution in MyGrid P.M.A. Sloot et al. (Eds.): EGC, LNCS 3470, pp. 444–453. Springer-Verlag Berlin Heidelberg.
- [2] Anjomshoaa, A., Brisard, F., Drescher, M., Fellows, D., Ly, A., McGough, S., Pulsipher, D. and Savva, A. (2005) Job Submission Description Language (JSDL) Specification, Version 1.0 (draft 19), Global Grid Forum 27.
- [3] Browne, J.C. (2004) "Grid Computing as Applied Distributed Computation: A Graduate Seminar on Internet and Grid Computing". IEEE international Symposium on Cluster Computing and the Grid, pp. 239-245.
- [4] Connolly, b. (1999) Distributed System and Database A Practical Approach to Design, Implementation, and Management. 2nd ed. England: Addison Wesley. pp. 603-638, 645-686.
- [5] Hoarau, W., and Tixeuil, S. (2008) Dependability Evaluation of the OGSA-DAI MiddleWare. Achievements in European Research on Grid Systems.
- [6] International Business Machines Corporation. OGSA-DAI Architecture. [Online]. [Accessed 17 March 2008]. Available from the World Wide

- Web:  
<http://www.globus.org/toolkit/docs/development/3.9.5/techpreview/ogsa-dai/doc/background/architecture.html>
- [7] Krause, A., Malaika, S., McCance, G., Magowan, J., Paton, N. And Riccardi, G. (2002) Grid Database. Service Specification,[Online]. Available from the World Wide Web: <http://www.cs.man.ac.uk/grid-db/papers/draft-ggf-dais-gdss-ggf8.pdf>.
- [8] Oliveria, C. H. P., Amaral, M. A., Oliveria, E.(n.d.) Algorithms for Distributed Database Query Command on the Grid. [Online]. [Accessed 28th May 2008]. Available from World Wide Web: <http://ccgrid07.lncc.br/docs/posters/27468.pdf>
- [9] Sidell, J. (1997) Performance of Adaptive Query Processing in the Mariposa Distributed Database Management System. [Online]. [Accessed 25th May 2008]. Available from World Wide Web: <http://www.db.cs.berkeley.edu/papers/unpubl-perf.pdf>.
- [10] Smith, J., Gounaris, A., Watson, P., Paton, N., Fernandes, A., Sakellarios, R. (2003) 'Distributed Query Processing on the Grid'. The International Journal of High Performance Computing Applications, vol. 17, pp. 353–367.
- [11] Stockinger, H. (2001) Distributed Database Management Systems and the Data Grid. Mass Storage Systems and Technologies. MSS '01. Eighteenth IEEE Symposium on, pp. 1-1.
- [12] Liebeherr, J., Akyildiz, I. F., and Omiecinski, E. (1991) Performance comparison of index partitioning schemes for distributed query processing. IEEE. In: System Sciences. Proceedings of the Twenty-Fourth Annual Hawaii International Conference, 8-11 Jan 1991, Kauai, HI, USA, 1, pp. 317-323.

# Impact of Guard Interval in Proposed MIMO-OFDM System for wireless communication

M.P.Chitra

Research Scholar, Sathyabama University,  
Chennai, India.  
chi\_mp2003@yahoo.co.in

Dr.S.K. Srivatsa

Senior Professor, St.Joseph College of Engineering,  
Chennai, India.  
profsks@rediffmail.com

**Abstract** - Alamouti's space-time coding scheme for Multi-Input Multi-Output (MIMO) system has drawn much attention in 4G wireless technologies. Orthogonal frequency division multiplexing (OFDM) is a popular method for high data rate wireless transmission. OFDM may be combined with antenna arrays at the transmitter and receiver to increase the diversity gain and enhance the system capacity on time variant and frequency selective channels, resulting in Multi-Input Multi-Output (MIMO) configuration. This paper explores various physical layer research challenges in MIMO-OFDM system design including channel modeling, space time block code techniques, channel estimation and signal processing algorithms used for performing time and frequency synchronization in MIMO-OFDM system. The proposed system is simulated in MATLAB and analyzed in terms of BER with signals to noise ratio (SNR). The difference of BER for coded and uncoded MIMO system and also the impact of guard interval are simulated using different wireless channel.

**Keywords** - Multi-Input Multi-Output (MIMO); orthogonal frequency division multiplexing (OFDM); Bit error rate (BER); signals to noise ratio (SNR); Single input single output (SISO); space time block code (STBC)

## I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) and space-time coding have been receiving increased attention due to their potential to provide increased capacity for next generation wireless systems. OFDM supports high data rate traffic by dividing the incoming serial data stream into parallel low-rate streams, which are simultaneously transmitted on orthogonal sub-carriers[1]. For large enough and a sufficiently large guard interval, the channels as seen by each of the sub-carriers become approximately frequency flat and allow for high order modulation. Due to this desirable feature, OFDM has been adopted in many commercial systems such as the IEEE 802.11a, ETSI HIPERLAN type2 wireless LAN systems and DAB, DVB-T broadcasting systems.

Space-time coding is a communication technique for wireless systems that realizes spatial diversity by introducing temporal and spatial correlation into the signals transmitted from different transmits antennas. Many space-time trellis and block codes have been proposed for flat fading channels. Most significantly, Alamouti discovered a very simple space-time block code (STBC) for transmission with two antennas guaranteeing full spatial diversity and full rate. It lends itself to

very simple decoding and has been adopted in third generation (3G) cellular systems such as W-CDMA. Recently, many literatures proposed space-time block coding schemes applicable to OFDM systems based on the Alamouti scheme [2]. When channel can be assumed to be approximately constant during two consecutive OFDM symbol durations, the Alamouti scheme is applied across two consecutive OFDM symbols and is referred to as the Alamouti STBC-OFDM or simply A-STBC-OFDM.

The combination of the multiple input multiple output (MIMO) signal processing with orthogonal frequency division multiplexing (OFDM) communication system is considered as a promising solution for enhancing the data rates of the next generation wireless communication systems operating in frequency selective fading environments. The High Throughput Task Group which establish IEEE 802.11n standard is going to draw up the next generation wireless local area network (WLAN) proposal based on the 802.11 a/g which is the current OFDM-based WLAN standards. The IEEE802.11n standard based on the MIMO OFDM system provides very high data throughput rate from the original data rate 54Mb/s to the rate in excess of 600 Mb/s because the technique of the MIMO can increase the data rate by extending an OFDM-based system. However, the IEEE 802.11n standard also increases the computational and hardware complexities greatly, compared with the current WLAN standards. It is a challenge to realize the physical layer of the MIMO OFDM system with minimal hardware complexity and power consumption.

The FFT/IFFT processor is one of the highest computational complexity modules in the physical layer of the IEEE 802.11n standard. If employing the traditional approach to solve the simultaneous multiple data sequences, several FFT processors are needed in the physical layer of a MIMO OFDM system. Thus the hardware complexity of the physical layer in MIMO OFDM system will be very high. This paper proposes an FFT processor with a novel multipath pipelined architecture to deal with the issue of the multiple data sequences for MIMO OFDM applications. The 128/64 FFT with 1-4 simultaneous data sequences can be supported in our proposed processor with minimal hardware complexity. Furthermore, the power consumption can also be saved by using higher radix FFT algorithm.

## II. CHANNEL MODELS

### A. Additive White Gaussian Noise channel

With the transmitted signal vector  $x$ , the received signal vector  $y$  is given by,  $y = x + n$  where 'n' represents additive white Gaussian noise vector. It follows the normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

$$f(n) = 1/\sqrt{2\pi\sigma^2} \exp(-(n-\mu)^2/2\sigma^2) \quad (1)$$

### B. Flat Fading channel model

It is modeled as,  $y = ax + n$  where  $a$  is the fading coefficients with PDF and  $n$  is the additive white Gaussian noise vector.

$$f(a) = 2a \exp(-a^2) \quad \text{for } a > 0. \quad (2)$$

### C. Frequency selective fading channel

In this model the channel is considered as a multi-path fading channel. It consists of multiple independent Rayleigh faders, which is modeled as complex-valued random processes. By assuming uniform antenna pattern and uniform distributed incident power, the received signal at the receiver can be expressed as

$$y = \sum_j a_j^* x + n \quad (3)$$

where 'n' is the additive white Gaussian noise and 'j' represents multi-path from transmitter.

## III. MIMO SYSTEM.

### A. Space – Time Codes.

Space-time codes (STC) provide transmits diversity for the Multi-Input Multi-Output fading channel. There are two main types of STC's namely space-time block codes (STBC) and space-time trellis codes (STTC). Space-time block codes operate on a block of input symbols, producing a matrix output whose columns represent time and rows represent antennas. Their main feature is the provision of full diversity with a very simple decoding scheme. On the other hand, Space-time trellis codes operate on one symbol at a time, producing a sequence of vector symbols whose length represents antennas. Like traditional TCM (Trellis Coded Modulation) for a single-antenna channel, Space-time trellis codes provide coding gain. Since they also provide full diversity gain, their key advantage over space-time block codes is the provision of coding gain [3]. Their disadvantage is that they are extremely hard to design and generally require high complexity encoders and decoders.

An STBC is defined by a  $p \times n$  transmission matrix  $G$ , whose entries are linear combinations of  $x_1, \dots, x_k$  and their conjugates  $x_1^*, \dots, x_k^*$ , and whose columns are pair wise –

orthogonal. When  $p = n$  and  $\{x_i\}$  are real,  $G$  is a linear processing orthogonal design which satisfies the condition that  $GTG = D$ , where  $D$  is the diagonal matrix with the  $(i,i)$ th diagonal element of the form  $(l_{1i}x_{12} + l_{2i}x_{22} + \dots + l_{ni}x_{n2})$ , with the coefficients  $l_{1i}, l_{2i}, \dots, l_{ni} > 0$ . Without loss of generality, the first row of  $G$  contains entries with positive signs. If not, one can always negate certain columns of  $G$  to arrive at a positive row.

$$G_2 = \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix} \quad G_4 = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \end{pmatrix} \quad (4)$$

We assume that transmission at the base-band employs a signal constellation  $A$  with  $2b$  elements. At the first time slot,  $nb$  bits arrive at the encoder and select constellation signals  $c_1, \dots, c_n$ . Setting  $x_i = c_i$  for  $i = 1, \dots, n$  in  $G$  yields a matrix  $C$  whose entries are linear combinations of the  $c_i$  and their conjugates. While  $G$  contains the  $n$  determinates  $x_1, \dots, x_n$ ,  $C$  contains specific  $c$  constellation symbols (or linear combinations of them), which are transmitted from the  $n$  antennas as follows: At time  $t$ , the entries of row  $t$  of  $C$  are simultaneously transmitted from the  $n$  antennas, with the  $i$ th antenna sending the  $i$ th entry of the row. So each row of  $C$  gives the symbols sent at a certain time, while each column of  $C$  gives the symbols sent by a certain antenna.

### B. Receive Diversity.

The base-band representation of the classical two-branch Maximal Ratio Receive Combining (MRRC) scheme. At a given time, a signal  $s_0$  is sent from the transmitter. The channel between the transmit antenna and the receive antenna zero is denoted by  $h_0$  and between the transmit antenna and the receive antenna one is denoted by  $h_1$  where  $h_0 = \alpha_0 e^{j\theta_0}$   $h_1 = \alpha_1 e^{j\theta_1}$ .

Noise and interference are added at the two receivers. The resulting received base band signals are  $r_0 = h_0 s_0 + n_0$ ,  $r_1 = h_1 s_1 + n_1$ .

Where  $n_0$  and  $n_1$  represent complex noise and interference.

Assuming  $n_0$  and  $n_1$  are Gaussian distributed, the maximum likelihood decision rule at the receiver for these received signals is to choose signal  $s_i$  if and only if (iff).

$$d^2(r_0, h_0 s_i) + d^2(r_1, h_1 s_i) \leq d^2(r_0, h_0 s_k) + d^2(r_1, h_1 s_k) \quad (5)$$

where  $d^2(x, y)$  is the squared Euclidean distance between signals  $x$  and  $y$  calculated by the following expression:

$$d^2(x, y) = (x - y)(x^* - y^*) \quad (6)$$

The receiver combining scheme for two-branch MRRC is as follows:



$$(\alpha_0^2 + \alpha_1^2 - 1)|s_i|^2 + d^2(s_0, s_i) \leq (\alpha_0^2 + \alpha_1^2 - 1)|s_k|^2 + d^2(s_0, s_k) \quad (7)$$

The maximal-ratio combiner may then construct the signal  $s_0'$ , so that the maximum likelihood detector may produce  $s_0''$ , which is a maximum likelihood estimate of  $s_0$ .

### C. Alamouti's Transmit Diversity Scheme.

#### 1) Two-Branch Transmit Diversity with One Receiver

The base-band representation of the two-branch transmit diversity scheme. The Encoding and Transmission Sequence at a given symbol period, two signals are simultaneously transmitted from the two antennas. The signal transmitted from antenna zero is denoted by  $s_0$  and from antenna one by  $s_1$ . During the next symbol period signal  $(-s_1^*)$  is transmitted from antenna zero, and signal  $s_0^*$  is transmitted from antenna one where  $*$  is the complex conjugate operation. The encoding is done in space and time (space-time coding) [4]. The encoding may also be done in space and frequency. Instead of two adjacent symbol periods, two adjacent carriers may be used (space-frequency).

TABLE I. ENCODING TABLE

	Antenna 0	Antenna 1
Time t	$s_0$	$s_1$
Time t+T	$-s_1^*$	$s_0^*$

#### 2) Transmit diversity with receiver diversity

It is possible to provide a diversity order of  $2M$  with two transmit and  $M$  receive antennas. For illustration, we discuss the special case of two transmit and two receive antennas in detail. The generalization to  $M$  receive antennas is trivial.

The base band representations of the scheme with two transmit and two receive antennas. The encoding and transmission sequence of the information symbols for this configuration is identical to the case of a single receiver.

Similarly, for  $s_1$ , using the decision rule is to choose signal  $s_i$  iff

$$(\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 - 1)|s_i|^2 + d^2(s_1, s_i) \leq (\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 - 1)|s_k|^2 + d^2(s_1, s_k) \quad (8)$$

The combined signals are equivalent to that of four branch MRRC,. Therefore, the resulting diversity order from the new two-branch transmit diversity scheme with two receivers is equal to that of the four-branch MRRC scheme.

It is interesting to note that the combined signals from the two receive antennas are the simple addition of the combined signals from each receive antenna. Hence conclude that, using two transmit and  $M$  receive antennas, using the combiner for each receive antenna and then simply add the combined signals from all the receive antennas to obtain the same diversity order as  $2M$ - branch MRRC.

### D. Channel Estimation.

#### 1) Enhance Channel Estimation

Frequency domain and is written in matrix notation

$$Y = SH + N \quad (9)$$

Where  $Y$  is the Fourier Transform of  $y$ ,  $S$  is the Fourier transforms of  $S$ ,  $N$  is the Fourier Transform of  $n$  and  $H$  is the Fourier transform of  $h$ .  $H$  can also be represented as

$$H = F.h \quad (10)$$

Where  $F$  is  $N \times N$  is the unitary FFT matrix. Therefore  $Y$  can be represented as,

$$Y = SF.h + N \quad (11)$$

$$Y = Qh + N \quad (12)$$

Where  $Q = X F$ . The estimated channel response in time domain can be obtained by the LS estimator as,

$$h = Q^H Q^{-1} Q^H Y \quad (12)$$

Where  $Q^H$  denotes the Hermitian transpose. The successful implementation of the estimator depends on the existence of the inverse matrix  $(Q^H Q)$ . If the matrix  $(Q^H Q)$  is singular (or close to singular), then the solution does not exist (or is not reliable) [5]. But it is a rare case.

### E. Training Sequence used.

To increase the performance of the channel estimation for OFDM systems in the presence of ISI, Kim and Stuber proposed this training sequence given by

$$X(n) = \begin{cases} A \exp(j2\pi(n/2)^2 / N) & n \in N \\ 0 & n \in M \end{cases} \quad (13)$$

where  $N$  is the set of sub-carrier odd indices, where  $M$  is the set of sub-carrier even indices.

Transmitted data with pilot. It has alternative zeros. By doing so, the transformation of the training sequence in the time domain has the special property that its first half is identical to its second half, while the desirable peak-to-average power ratio of one is still retained. In our work, this training sequence is applied to the LS estimator for MIMO-OFDM systems.

### F. Channel coefficients.

The Actual, estimated coefficients through least square estimator and error between them. These Coefficients are

generated using Monte- carlo simulation. The error is in the order of 10-3.

TABLE II. CHANNEL COEFFICIENTS

Estimated	Actual	Error
-0.7239 - 0.6893i	-0.7243 + 0.6895i	-0.0004
-0.0626 - 0.6063i	-0.0627 + 0.6063i	-0.0000
-0.1315 + 0.4757i	-0.1317 - 0.4766i	-0.0009
-0.3951 - 0.0034i	-0.3940 + 0.0030i	0.0011
0.0143 + 0.2363i	0.0138 - 0.2367i	-0.0004
-0.1753 + 0.0735i	-0.1752 - 0.0735i	0.0001
0.1065 + 0.0430i	0.1077 - 0.0429i	-0.0011
-0.0655 + 0.0239i	-0.0652 - 0.0252i	-0.0002
0.0411 + 0.0211i	0.0412 - 0.0209i	0.0000

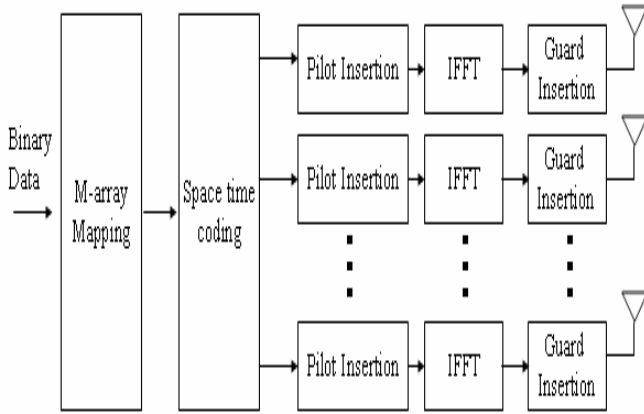


Figure 1. Transmitter

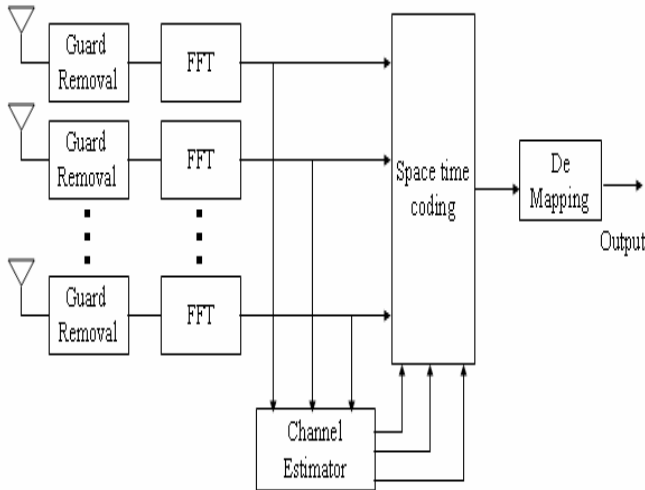


Figure 2. Receiver

#### IV. MIMO – OFDM SYSTEM

In the area of Wireless communications, MIMO-OFDM is considered as a mature and well establishes technology. The main advantage is that it allows transmission over highly

frequency-selective channels at a reduced Bit Error Rate (BER) with high quality signal. One of the most important properties of OFDM transmissions is the robustness against multi-path delay spread [6]. This is achieved by having a long symbol period, which minimizes the inter-symbol interference. Unfortunately, this condition is difficult to fulfill in MIMO-OFDM systems, since the GI length is a system parameter, which is assigned by the transmitter. But the maximum propagation delay is a parameter of the channel, which depends on the transmission environment. MIMO can be used either for improving the SNR or data rate. For improving the data rate, A-STBC-OFDM system is used.

##### A. Proposed FFT Algorithm

Given a sequences  $x(n)$ , an  $N$ -points discrete Fourier transform (DFT) is defined as

$$X(k) = \sum_{n=0}^{N-1} x[n] W_N^{kn} \quad (14)$$

$$k=0,1,\dots,127.$$

Where  $x[n]$  and  $X(k)$  are complex numbers. The twiddle factor is

$$W_N^{nk} = e^{-j(2\pi nk/N)} = \cos\left(\frac{2\pi nk}{N}\right) - j \sin\left(\frac{2\pi nk}{N}\right) \quad (15)$$

Because 128- point FFT is not a power of 8, the mixed – radix FFT algorithm, including the radix-2 and radix -8 FFT algorithms, is needed. Since the algorithm has been derived in detail previously [7], it will be described briefly here

First let

$$N=128$$

$$n=64n_1+n_2, \{n_1=0,1 \text{ and } n_2=0,1,\dots,63.$$

$$k=k_1+2k_2, \{k_1=0,1 \text{ and } k_2=0,1,\dots,63. \quad (16)$$

Using (16),(14) can be rewritten as,

$$X(2k_2 + k_1) = \sum_{n_2=0}^{63} \sum_{n_1=0}^1 x[64n_1 + n_2] W_{128}^{(64n_1+n_2)(2k_2+k_1)}$$

$$= \sum_{n_2=0}^{63} \left\{ \sum_{n_1=0}^1 x(64n_1 + n_2) W_2^{n_1 k_1} W_{128}^{n_2 k_1} \right\} W_{64}^{n_2 k_2} \quad (17)$$

## V. SIMULATION RESULTS

TABLE III. SIMULATION PARAMETERS

Parameter	Specification
Number of Sub-carrier	64
FFT size	64
Modulation type	BPSK
Channel model	AWGN, Fading Channel
Doppler Frequency	50Hz
Guard Interval	10
Guard Interval	10

The performance of SISO systems under AWGN and Fading channel. From the graph, the following observations are made. In additive white Gaussian noise (AWGN), using typical modulation and coding schemes, reducing the effective bit error rate (BER) from  $10^{-2}$  to  $10^{-3}$  may require only 2 or 3 dB higher signal to-noise ratio (SNR). Achieving the same in a multi-path fading environment, however, may require up to 10 dB improvement in SNR.

In fading channel, using typical modulation and coding schemes, reducing the effective bit error rate (BER) in MIMO systems from  $10^{-2}$  to  $10^{-3}$  may require only 1-4 dB SNR. Achieving the same in SISO system required greater than 10 dB SNR.

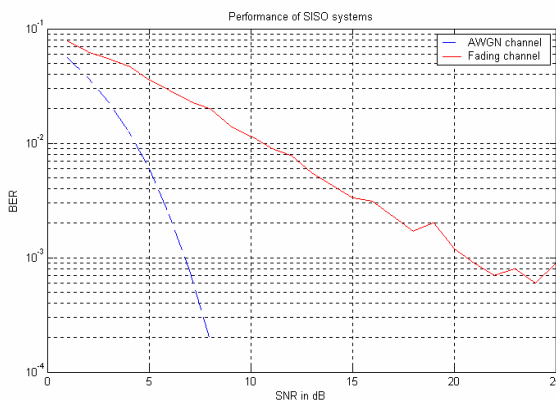


Figure 3. Performance of SISO Systems.

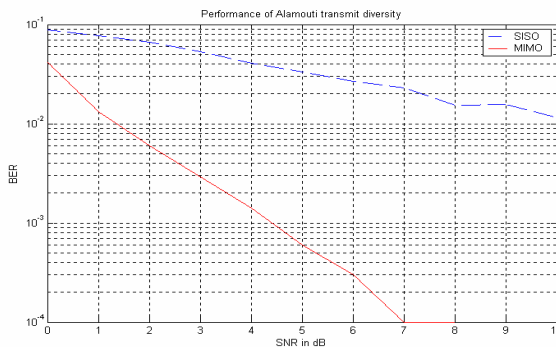


Figure 4. Performance of Alamouti's transmit diversity

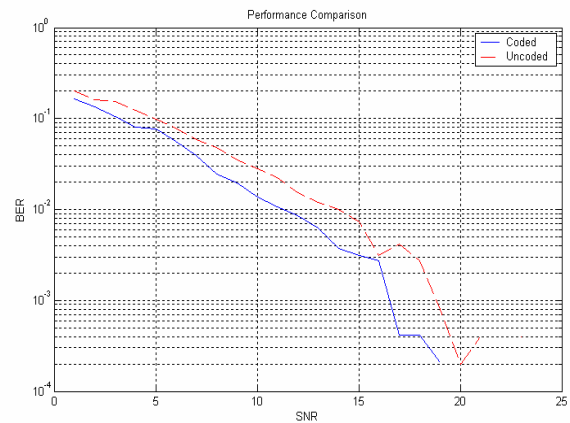


Figure 5. Performance of A-STBC – OFDM with MIMO – OFDM.

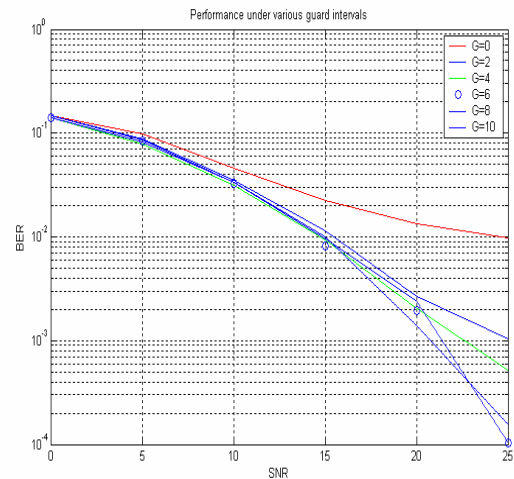


Figure 6. Performance Under various Guard Interval

The performance of A-STBC OFDM with MIMO-OFDM, obviously space time coded system performs well in higher SNR region when the SNR is greater than 15 dB the BER is less than  $10^{-3}$  in coded MIMO-OFDM system. But uncoded MIMO system the bit error rate is greater than  $10^{-2}$  when the SNR is greater than 15 dB.

The performance of the MIMO-OFDM is increased when the guard interval is increased. When the guard interval is 10, the BER is decreased less than  $10^{-2}$  in SNR 15dB.

## CONCLUSION

OFDM is an effective technique to combat multi-path delay spread for wideband wireless transmission. OFDM with multiple transmit and receive antennas form a MIMO system to increase system capacity. The system with STC (A-STBC-OFDM) and with high guard interval achieves the system requirements of high quality transmission and high data rate transmission. The performance of the MIMO – OFDM system is optimized with minimum bit error rate.

REFERENCES

- [1] I. Barhum, G. Leus, and M. Moonen, "Optimal training design for mimo ofdm systems in mobile wireless channels," *IEEE Trans. Signal Processing*, vol. 5, pp. 1615–1624, June 2003.
- [2] Die Hu, Luxi Yang, Yuhui Shi, and Lianghua He, "Optimal Pilot Sequence Design for Channel Estimation in MIMO OFDM Systems" *IEEE COMMUNICATIONS LETTERS*, VOL. 10, NO. 1, JANUARY 2006.
- [3] Sarod Yatawatta and Athina P. Petropulu, "Blind Channel Estimation in MIMO OFDM Systems with Multiuser Interference" *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, VOL. 54, NO. 3, MARCH 2006.
- [4] Nadem H. Dawod, Ian D. Marsland, and Roshdy H. M. Hafez, "Improved Transmit Null Steering for MIMO – OFDM Downlinks With Distributed Base Station Antenna Arrays" *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, VOL. 24, NO. 3, MARCH 2006.
- [5] Tsuyoshi Kashima, Kazuhiko Fukawa, and Hiroshi Suzuki, "Adaptive MAP Receiver via the EM Algorithm and Message Passings for MIMO-OFDM Mobile Communications" *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, VOL. 24, NO. 3, MARCH 2006.
- [6] Heunchul Lee, Byeonsi Lee and Inkyu Lee. "Iterative Detection and Decoding With an Improved V-BLAST for MIMO-OFDM Systems" *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, VOL. 24, NO. 3 MARCH 2006.

# Implementation of New Routing Protocol for Node Security in a Mobile Ad Hoc Network

Virendra Singh Kushwah

Department of Computer Science

Hindustan Institute of Management and Computer Studies

Farah, Mathura, INDIA

kushwah.virendra248@gmail.com

Gaurav Sharma

Department of Computer Science

GLA, University

Mathura, INDIA

gauravsharma53@gmail.com

**Abstract--** A routing protocol plays important role to handle entire network for communication and determines the paths of packets. A node is a part of the defined network for transferring information in form of packets. If all packets transferred from source to destination successfully, it has been assumed that the routing protocol is good. But, an attacker turns this dealing as a speed breaker and turning point of a highway. So, prevention from attacks and secure packets, a new routing protocol is being introduced in this paper. The proposed routing protocol is called by SNAODV (Secure Node AODV). This paper is also tried to maximize throughput as compared with AODV and SAODV.

**Keywords—**AODV; routing; packets; network

## I. INTRODUCTION

A mobile ad hoc network (MANET) consists of a group of devices (or nodes) that rely on the wireless communication medium and themselves for data transmission. A node in an ad hoc network has direct connection with a set of nodes, called neighbouring nodes, which are in its communication range. The number of nodes in the network is not necessarily fixed. A MANET does not have base stations or routers. Each node acts as a router and is responsible for dynamically discovering other nodes it can directly communicate with. However, when a message without encryption is sent out through a general tunnel, it may be maliciously attacked. Nodes cooperate by forwarding packets on behalf of each other when destinations are out of their direct wireless transmission range. A centralized administrator and/or a pre-deployed network infrastructure are not necessary for a MANET to be set up, thus making its deployment quick and inexpensive. In addition, Nodes ability to move freely ensures a flexible and versatile dynamic network topology which can be desirable in many situations. Hence, in addition to acting as hosts, each mobile node does the functioning of routing and relaying messages for other mobile nodes. Being mobile, any node can communicate to other nodes. Nodes do not necessarily know each other and come together to form an ad hoc group for some specific purpose. While limited bandwidth, memory, processing capabilities and open medium make its disadvantages. There are two types of possible attacks on nodes in MANET: passive attacks and active attacks. In passive attacks, adversaries simply drop and refuse to forward other nodes requests of assigning keys. In active attacks, in contrast, adversaries may

return a fake reply (e.g. an invalid partial key) to the node requesting key. However, the security of MANET is still a challenge issue.

## II. PROBLEM STATEMENT

There are a number of solutions for securing routing protocols in MANETs. We know there are two authentication models for securing routing is available that are ARAN [14] and SAODV [15] since they are closely related to our proposed model. In general, the existing schemas/models for secure routing are based on the assumptions of the availability of key management infrastructures which are unrealistic and contrast to the ad hoc network concepts. Moreover, these schemas do not consider intermediate nodes during the routing steps; therefore, the nodes may perform fabrication attacks. From these weaknesses of current approaches, our goal is to design a schema which performs point-to-point message authentication without a deployed key management infrastructure.

When two nodes are communicating, there may be any chance to steal packets, destroy packets or corrupt packets by malicious nodes. There are following two questions:-

1. Are nodes making right communication?
2. Are packets being saved during transmissions?

If these two questions are solved, at least it is understandable to prevent from misbehaviour nodes which make interfered between two or more right nodes during transmission of packets. So prevention is better than cure. To detect malicious nodes and remove those nodes is two way process [2]. So follow two processes, it is better process to use certificate on those nodes. If those nodes are secured, at least packets can be saved from attackers during transmission.

## III. LITERATURES REVIEW

Security has become wide research area in MANETs. Most existing papers on deploying key management in MANETs usually mention flooding briefly as a way to distribute key in an ad hoc network using AODV routing protocol. Most secure communication protocols rely on a secure, robust and efficient key management scheme. Key management is also a central aspect for security in mobile ad hoc networks. In mobile ad



hoc networks, the computational load and complexity for key management are strongly subject to restriction by the node's available resources and the dynamic nature of network topology.

- A. A secure identity based key management scheme is proposed suitable for applying in MANETs. Similar to other ID-based cryptosystems, a trusted key generation center is needed in this scheme for verifying user identity and generating the corresponding private keys. [4]
- B. Research work in key management scheme and handlings about limited number of nodes are possible in an ad hoc network. When the number of nodes increases, most of them become either inefficient or insecure. The main problem of any public key based security system is to make user's public key available to others in such a way that is authenticity is verifiable. [5]
- C. Using novel hierarchical security scheme, called Autonomous Key Management (AKM), which can achieve flexibility and adaptivity, and handles MANET with a large number of nodes. AKM is based on the hierarchical structure and secret sharing to distribute cryptographic keys and provide certification services. AKM also enables the ability to issue certificates with different levels of assurance with the help of a relatively small number of nodes. [6]
- D. SEKM (Secure and Efficient Key Management) builds a public key infrastructure (PKI) by applying a secret sharing scheme and using an underlying multicast server groups. In SEKM, each server group creates a view of the certificate authority (CA) and provides certificate update service for all nodes, including the servers themselves. The advantage is that in SEKM it is easier for a node to request service from a well maintained group rather than from multiple "independent" service providers which may be spread in a whole area. [7]
- E. In Authenticated Acknowledgement Scheme (AAS) to detect such selfish nodes, routes containing such nodes will be eliminated from consideration. The source node will be able to choose an appropriate route to send its data. The AAS scheme is a network-layer technique to detect the selfish nodes and to mitigate their effects. It can be implemented as an add-on to existing routing protocols for MANETs, such as DSR. The AAS scheme detects misbehavior through the use of a new type of authenticated acknowledgment scheme termed AAS, which assigns a fixed route of two hops (three nodes) in the opposite direction of the data traffic route. When a node wishes to communicate with another node, a methodology is performed by the sending and receiving nodes, which ensures authentication and integrity. [8]
- F. In [14], the authors categorized three kinds of threats which are modification, impersonation and fabrication in AODV and DSR. On the basic of this analysis, the authors proposed a protocol called ARAN (Authenticated Routing for Ad hoc Networks) using cryptographic certificates to bring authentication, message-integrity and non-repudiation to the route discovery process based on the assumption of existing of a trusted certificate server. It is not appropriate with ad hoc networks because it forms a centralized element. Moreover, in this protocol, because the source node cannot authenticate intermediate nodes in the routing path, intermediate malicious nodes can use error message attacks to networks.
- G. In [15], the authors extend the AODV routing protocol to guarantee security based on the approach of key management scheme in which each node must have certificated public keys of all nodes in the network. This work uses two mechanisms to secure the AODV messages: digital signature to authenticate the fixed fields of the messages and hash chains to secure the hop count field. This protocol uses public key distribution approach in the ad hoc network; therefore, it is difficult to deploy and computationally heavy since it requires both asymmetric cryptography and hash chains in exchanging messages. The protocol also did not consider the authentication of intermediate nodes; hence it could not prevent the attack of falsifying error messages in ad hoc networks.
- H. The authentication mechanism is based on using the IEEE 802.11i standard, Extensible Authentication Protocol (EAP) and RADIUS (Remote Authentication Dial in User Service). The proposed authentication system is divided in two mechanisms: Level 1 and Level 2, and it are a combination of a centralized and a distributed system. The proposed solution is a 2-Level Authentication mechanism. The first level (Level 1) is used for authenticating new nodes that request access to the MANET. Level 1 authentication is used when the MANET does not have any connection to other infrastructure (e.g. the Internet), and therefore has no centralized authentication server. However, the MANET could have a centralized key management service. This is determined by the security policy. The Level 2 authentication process is used when the MANET is connected to the Internet, with a centralized authentication server (RADIUS). [9]

#### IV. SYSTEM MODEL

The principle of our model is that messages in ad hoc network must be authenticated to guarantee the integrity and non-repudiation so that the protocol and nodes can be prevented against several kinds of attacks. Each node in a network has its own a pair of public key  $e$  and private key  $d$  following RSA Public-key Crypto-system [13] by self-generation, and each node contains a list of neighbour nodes with records containing the information of a neighbour node including neighbour address, neighbour public key, and a shared secret key. This information is formed after the key agreement between two neighbour nodes to negotiate a pair of keys and a shared secret key. The details of our security schema for AODV are described as the following sections.

##### A. Key Agreement Process between Neighbor Nodes

A node joining a network requires sending key agreement messages to its neighbours to negotiate a shared secret key. The concept of this process is based on HELLO message in ad-hoc routing protocols. The node broadcasts a message indicating the negotiation request with neighbour nodes:

$\langle \text{KEY\_AGREEMENT\_REQ}, \text{request\_id}, \text{sender\_address}, \text{PK}_S \rangle$

On receiving this request, nodes reply a message:

$\langle \text{KEY\_AGREEMENT\_REP}, \text{request\_id}, \text{sender\_address}, \text{neighbour\_address}, \text{PK}_N \rangle$

(Where  $\text{PK}_S$  and  $\text{PK}_N$  are the public key of the sender node and replying node, respectively;  $\text{request\_id}$  is a sequence number generated by the sender node) to indicate the receiving of the request message and inform that it is ready for the key agreement process. For each received message, the request node (i.e.; node A) creates a new record in its neighbour list. Each record contains filled neighbour address and filled neighbour public key; the other fields of the record are empty. For each new record in the list, the request node (A) negotiates a secret key with the neighbour node (B) by the following steps:

1. Generate a key  $K_s$  by using a secure random number generator,
2. Encrypt  $K_s$  with  $\text{PK}_B$  (node B's public key) = encrypt  $\text{PK}_B(K_s)$ ,
3. Send an offer message  
 $\langle \text{KEY\_PASS}, \text{encrypt } \text{PK}_B(K_s) \rangle$  to B,
4. Wait ACK (acknowledgement) from B and check message integrity to finish the negotiation

When node B receives the key passing message, it decrypts " $\text{encrypt } \text{PK}_B(K_s)$ " by its private key ( $\text{pB}$ ) to get the shared key  $K$ . Then, node B sends the ACK message

$\langle \text{KEY\_PASS\_ACK}, \text{request\_id}, \text{HASH}_{K_s}(\text{request\_id}) \rangle$

to indicate successful shared secret key negotiation, where  $\text{HASH}_{K_s}(\text{request\_id})$  is the hashed message of  $\text{request\_id}$  by the shared key  $K_s$ .

Since RSA algorithm is used in the negotiation, the confidentiality of the shared key is guaranteed between the two nodes. The shared key is used for authenticating messages between two adjacent nodes later in AODV routing protocol. In the case a node does not have a shared key with its neighbour nodes; it cannot participate in routing transactions.

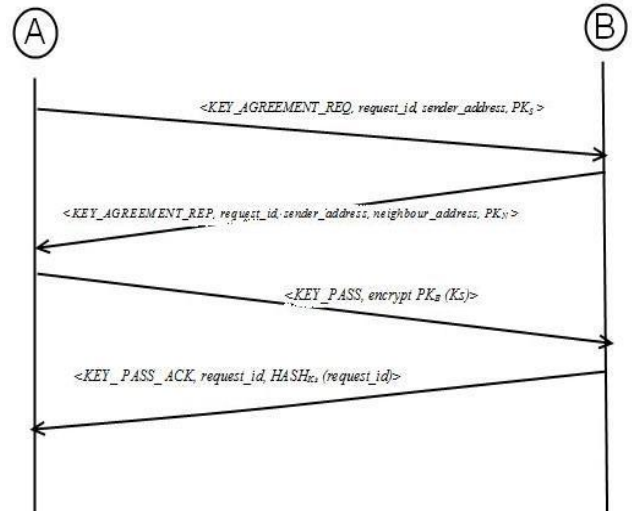


Figure 1. Node to node authentication process

##### B. Route Request

Route request (RREQ) is initiated by a source node (S) and then propagated by intermediate nodes until the message reaches its destination node (D). On receiving RREQ, an intermediate node I, according to our designed routing protocol, checks whether the message will be re-broadcasted or not. If the message needs to be re-broadcasted and the sender is in node I's neighbour list, it will send (unicast) a message to request the authentication process from the sender:

$\langle \text{RREQ\_REQ}, \text{source\_address}, \text{broadcast\_id} \rangle$ .

When receiving the authentication request, the sender creates an authentication reply message containing

$\langle \text{RREQ\_REP}, \text{source\_address}, \text{broadcast\_id}, \text{HASH}_{K_s}(\text{RREQ}) \rangle$

Where  $\text{HASH}_{K_s}(\text{RREQ})$  is the hashed value of RREQ message by the shared key  $K_s$  between the two nodes. The authentication reply message is unicasted back to node I. Node I on receiving the message will check the integrity of the RREQ message by hashing the message with using the shared

key  $K_s$  and then comparing with the received hashed digest. If the comparison is successful (the integrity of the RREQ message is guaranteed), node I continues steps following AODV such as set up reverse path, increase the hop count, rebroadcast the message and so on; otherwise, the RREQ will be discarded. The process continues until the message reaches the destination. The destination also authenticates the sender of RREQ (neighbour of the destination) by the same procedure.

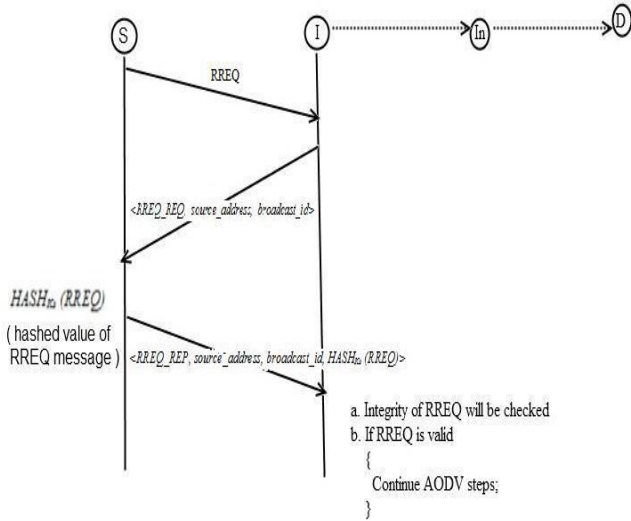


Figure 2. Representation of the message authentication

### C. Route Reply and Route Maintenance

Route replies (RREP) in AODV are also targets for attacks by malicious nodes. In our model, when receiving a RREP, a node requests the sender to prove the integrity and non-repudiation of the message by sending an authentication message. The request for authentication is

$\langle RREP\_REQ, destination\_address, destination\_sequence\# \rangle$

and the reply is

$\langle RREP\_REP, destination\_address, destination\_sequence\#, HASH_{K_s}(RREP) \rangle$

where  $HASH_{K_s}(RREP)$  is the hashed value of RREP message by the shared key  $K_s$  between the two nodes. After the authentication process is successful, a node continues to the steps in AODV, otherwise, the node drops RREP since it is invalid.

In route maintenance process, only route error report message (RERR) is a target for attacks in AODV protocol. Our schema requires the authentication process in sending route error

messages to prevent attacks from malicious nodes. The authentication request and response for RERR is

$\langle RERR\_REQ, unreachable\_destination\_address, unreachable\_destination\_sequence\# \rangle$ ,

and

$\langle RERR\_REP, unreachable\_destination\_address, unreachable\_destination\_sequence\#, HASH_{K_s}(RERR) \rangle$ , respectively.

### D. Routing Message formats

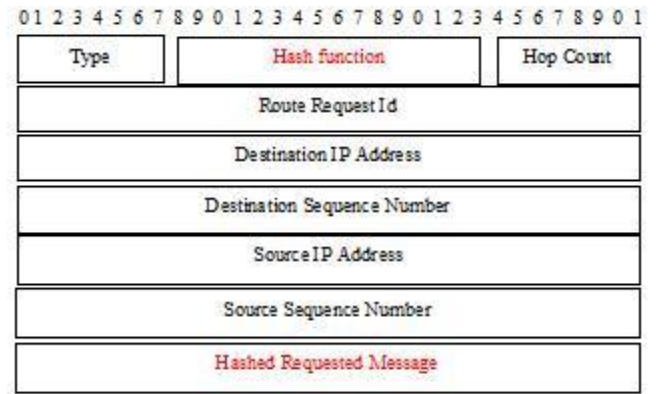


Figure 3. RREQ message format of SNAODV

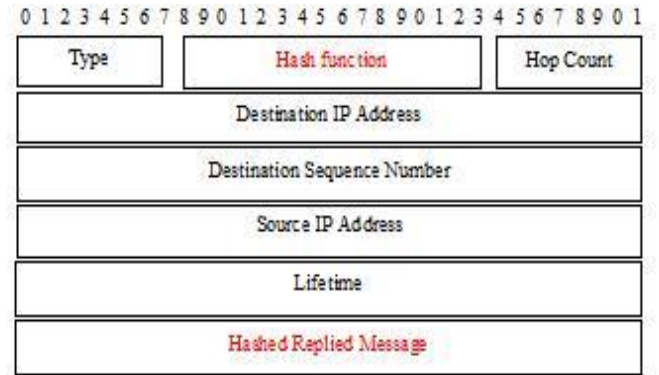


Figure 4. RREP message format of SNAODV

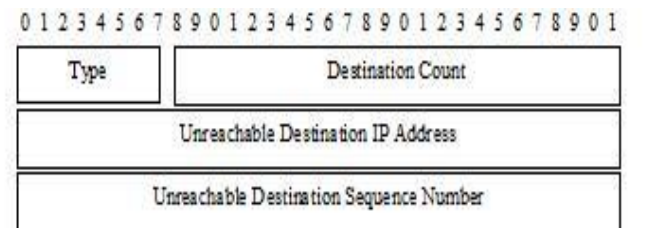


Figure 5. RERR message format of SNAODV

E. Algorithm for node-to-node authentication of the System Model



1. Sender node broadcasts a message indicating the negotiation request with neighbour nodes

$\langle \text{KEY\_AGREEMENT\_REQ}, \text{request\_id}, \text{sender\_address}, \text{PK}_S \rangle$

2. Sender node gets reply a message

$\langle \text{KEY\_AGREEMENT\_REP}, \text{request\_id}, \text{sender\_address}, \text{neighbour\_address}, \text{PK}_N \rangle$

3. The request node (A) negotiates a secret key with the neighbour node (B) by the following steps:

- a. Generate a key  $K_s$  by using a secure random number generator,
- b. Encrypt  $K_s$  with  $\text{PK}_B$  (node B's public key) = encrypt  $\text{PK}_B(K_s)$ ,
- c. Send an offer message

$\langle \text{KEY\_PASS}, \text{encrypt } \text{PK}_B(K_s) \rangle$  to B,

- d. Wait ACK (acknowledgement) from B and check message integrity to finish the negotiation
4. Node B sends the ACK message

$\langle \text{KEY\_PASS\_ACK}, \text{request\_id}, \text{HASH}_{K_s}(\text{request\_id}) \rangle$

## V. SIMULATION AND RESULTS

Simulation of the work has been done on Qualnet 5.0.1 for implementing new designed routing protocol. We have implemented RREQ and RREP message formats for new routing protocol using hash function i.e.; MD5 (Message Digest 5). It has been given in above figures. Simulation done on the following parameters basis:

TABLE I. SIMULATION SETUP

Parameters	Value
Simulation Area	1500m x 1500m
Number of nodes	30 (4 nodes are wormhole)
Simulation duration	120 s
Routing protocol	AODV and SNAODV
Mobility pattern of nodes	Random waypoint

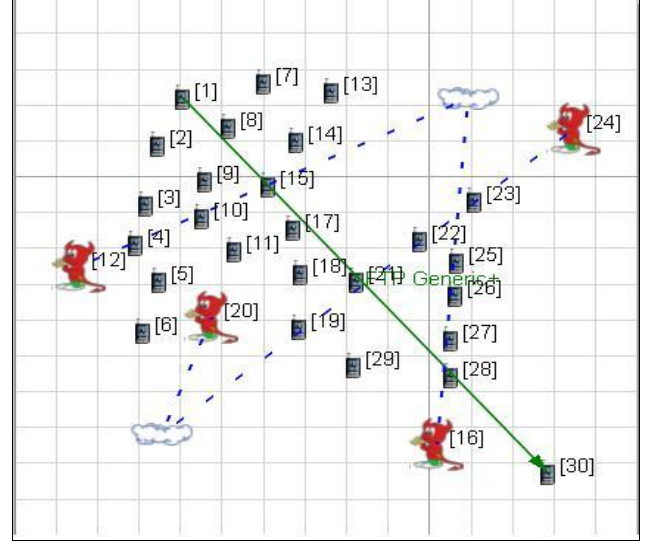


Figure 6. 30 nodes MANET environment with 4 black hole nodes

TABLE II. SIMULATION RESULTS

Parameters	AODV	SNAODV
Throughput	2435	2700
Number of RREQ packets initiated	19	24
Number of data packets sent as source	183	208
Number of data packets received	62	64
Number of RREQ packets retried	30	29
Number of RREQ packets received by dest	19	23
Number of RREP packets initiated as dest	14	20
Number of RREP packets received as Source	15	22
Number of Data Packets Dropped for no route	Node1=31, Node2 = 1, Node11 = 1	Node1=24, Node2 = 11, Node11 = 0

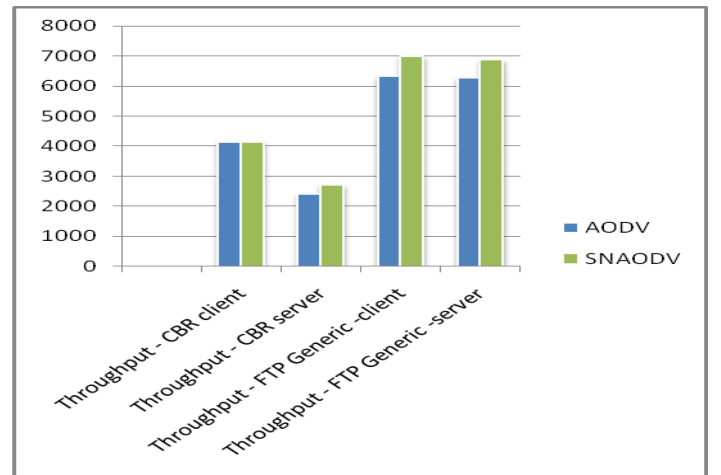


Figure 7. Throughput based comparison between AODV & SNAODV

The results have been come out from simulated on Qualnet 5.0 tool on the above simulation parameters and the results are being shown that the goal of new protocol to maximize the throughput. Throughput values of CBR client of both routing protocols are same while throughput values of CBR server is different in our new proposed protocol has higher values than AODV. Same process is in FTP Generic server.

## VI. CONCLUSION

This paper presents a new secure routing protocol for MANETs. It also provides node to node authentication and enables mobile user to ensure the authenticity of user of peer node. The significant advantage of our solution is to get all packets meaning that packet will be transmitted from source to destination without losing packets. The system model solved the security problem in the ad hoc network and is also suitable for application to other wired and wireless network. This paper is maximizing throughput of the network on the various parameters. One advantage of the SNAODV protocol is that no key assumption is required like SAODV has.

## REFERENCES

- [1] L.Zhou and Z.Haas, "Securing AdHoc Networks," IEEE Network, vol.13, no.6, page no.24–30, November/December 1999
- [2] B. Sukla, "Detection/Removal of Cooperative Black and Gray Hole Attack in Mobile Ad-Hoc Networks", In proceeding of the World Congress on Engineering and Computer Science 2008, October 22-24,2008,San Francisco,USA
- [3] Nguyen H and Morino H, "A Key Management Scheme for Mobile Ad Hoc Networks Based on Threshold Cryptography for Providing Fast Authentication and Low Signaling Load", IFIP International Federation for Information Processing 2005, LNCS 3823, page no. 905-915,2005
- [4] A.Kapil and S.Rana, "Identity-Based Key Management in MANETs using Public Key Cryptography", International Journal of Security (IJS), Volume (3): Issue (1), published in Computer Science journal in March, 2009.
- [5] S. Capkun, L. Buttyan, J.-P. Hubaux, "Self-organized public-key management for mobile ad hoc networks", Technical Report 2002/34, EPFL/IC, May 2002.
- [6] Bo. Zhu, Robert H. Deng, Mohan S. Kankanhalli, Guilin Wang, "Efficient and robust key management for large mobile ad hoc networks", In Proceedings of Computer Networks 48 (2005), page no. 657–682, 2005.
- [7] Bing Wu, Jie Wu, Eduardo B. Fernandez, Mohammad Ilyas, Spyros Magliveras, "Secure and efficient key management in mobile ad hoc networks", Journal of Network and Computer Applications 30 (2007), page no. 937–954, 2007.
- [8] M. Gunasekaran, P. Sampath and B. Gopalakrishnan, "AAS: An Authenticated Acknowledgement-Based Scheme for Preventing Selfish Nodes in Mobile Ad Hoc Networks", International Journal of Recent Trends in Engineering, Vol 1, No. 1, May 2009, page no. 294-298, 2009.

- [9] Andreas Hafslund and Jon Andersson, Thales Norway AS, "2-Level Authentication Mechanism in an Internet connected MANET", 6<sup>th</sup> Scandinavian Workshop on Wireless Ad-hoc Networks, 2006.
- [10] Marcelo M. Carvalho, Cintia B. Margi, Katia Obraczka and J. J. Garcia-Luna-Aceves, "Modeling Energy Consumption in Single-Hop IEEE 802.11 Ad Hoc Networks", Proceeding in 13<sup>th</sup> International conference on Computer communication and networks 2004, 2004.
- [11] Qualnet 4.5.1 Network Security model library.pdf
- [12] K. Sanzgiri, D. Laflamme, B. Dahill, B. Levine, C. Shields and E. Royer. An Authenticated Routing for Secure Ad Hoc Networks. Journal on Selected Areas in Communications special issue on Wireless Ad hoc Networks, March 2005.
- [13] Man,Y.R.: Internet Security cryptographic: principles, algorithms and protocols. Wiley Publishing House, Chichester(2004).
- [14] Kimaya, S., et al.: Authenticated routing for ad hoc networks. Journal on Selected Areas in Communications 23(3), 598–610 (2005).
- [15] Zapata, M.G., Asokan, and N.: Securing Ad hoc Routing Protocols. In: Proc. of the ACM workshop on Wireless security, Atlanta, USA, pp. 1–10 (2002)

## AUTHORS PROFILE



Mr. Virendra Singh Kushwah is an Assistant Professor of Computer Science, Hindustan Institute of Management and Computer Studies, Farah, Mathura, INDIA. He received his M.Tech from ABV-Indian Institute of Information Technology and Management, Gwalior, INDIA. His research areas are Information Security, Ad hoc network and Image Processing. He has published his research papers in International Journals and Conferences including IEEE etc.



Mr. Gaurav Sharma is an Assistant Professor of Computer Science, GLA University, Mathura, INDIA. He received his MCA from UPTU and presently, he is pursuing in PhD in Computer Science and Engineering. He has more 5 years experience. His research areas are Information Security, Computer Network etc.



# Runtime Monitoring and controlling of Information flow

Mohamed Sarrab

Software Technology Research Laboratory  
De Montfort University  
Leicester, LE1 9BH UK  
msarrab@dmu.ac.uk

Helge Janicke

Software Technology Research Laboratory  
De Montfort University  
Leicester, LE1 9BH UK  
heljanic@dmu.ac.uk

**Abstract—** Computer systems are monitored to check performance or verified to check the correctness of the software systems with respect to security properties such as integrity, availability and confidentiality. The more sensitive the information, such as credit card information, military information or personal medical information, being processed by the software, the more necessary and important it is to monitor and control the flow of sensitive information. Monitoring and controlling an untrusted program behavior to ensure information confidentiality during runtime in an environment where confidential information is present is often difficult and unnerving task for users. The issue is how users can monitor and control the flow of the confidential information at the runtime of untrusted programs. In this paper we present a novel runtime verification approach for monitoring and controlling information flow that supports user interaction with a running program.

**Keywords-**Information flow control; Runtime monitoring; Confidentiality; Usability.

## I. INTRODUCTION

As our businesses, government and military become increasingly dependent on modern information technology, computer application security protection against malicious code and software system bugs become increasingly important. The more sensitive the information, such as credit card data, government intelligence, military or personal medical information being processed by software, the more important it is to ensure this information confidentiality. The leakage of confidential information may cause financial damage in case of loss or destroy private or sensitive secret information. As an example Trusted Solaris [1] uses a security technique that determines which information is accessible by users, using a mandatory access control mechanism. However, in many cases discretionary access mechanism that is usable, reliable and can protect the confidentiality and integrity of sensitive information accessed by any untrusted software are more suitable as they do not involve the source level of administration and grant users discretion about how their information is being used. Information flow occurs from a source (subject) to a target or destination (object) whenever information stored in a source is propagated directly or indirectly to a target object. An example flow would be the

copying of a file into an email that is subsequently sent through the network. The following informal examples illustrate this:

Assuming that some sensitive information is stored on a computer system, how can we prevent it from being leaked? The first approach that comes to mind is to limit access to this sensitive information, by using any type of access control mechanisms or using encryption or firewalls mechanisms. These are very useful approaches which however have their limitations. Standard security mechanisms are focused only on controlling the release of information but no restrictions are placed on the propagation of that information and thus are unsatisfactory for protecting confidential information.

Standard security mechanism Such as Access control, Firewall and Encryption [2, 24] only focus on controlling the release of information but no limitations are placed on controlling the propagation of that confidential information. There is no monitoring mechanism for controlling information flow during runtime that support:

- **Configurable information flow** Configurability is an important requirement because what is considered to be secure today can be insecure tomorrow. A property of configurable information flow is that it provides flexible security mechanisms that can control changeable security requirements.
- **User interaction control** Interaction with users is very important in flexible and reliable security monitoring mechanism because different users may have different security requirements. That cannot always be intercepted prior to the execution of the program. Users interact with a monitoring mechanism during runtime, enabling them to change program behaviors or modify the way that information flows.

The main objective of this research is to develop a usable security mechanism for controlling information flow within a software application during runtime. Usable security have refers to enabling users to manage their systems security without defining elaborate security rules before starting the application. Security will be achieved by an interactive process in which our framework will query the user for security requirements for specific information that are made available to the software and then continue to enforce these

requirements on the application using runtime verification technique for tracing information flow. There are several research problems associated with these objectives, that we are addressing in this paper.

**Policy Specification:** The gap between user understandable information flow policies and user security requirements. Security requirements are often expressed in natural language (English) and in many cases carry ambiguity. Translating requirements into formal unambiguous policy specification is difficult.

**Policy configuration:** The user must be able to modify the information flow policies during runtime. Policies are dynamic and can change in response to user interaction. Dynamic policies and reconfiguration of their enforcement mechanisms are not well understood.

**Runtime Monitoring:** The monitoring mechanism ensures that the program contains only legal flows. Traditional runtime monitoring only addresses the monitoring of safety properties. They are not suitable for monitoring information flow or managing the program behavior at runtime, as there is no feedback from the monitor to the observed software.

**Runtime Management:** The behavior of a program leaking confidential information will be altered by the monitor according to the user decision. Analyzing the impact of a user or policy induced program alteration with the program original functional requirements is an open question.

**A dynamic policy based approach:** A dynamic property provides a flexible and usable security mechanism that can control changeable security requirements. Dynamic policy based approach has been before [3] but only for access control and not for information flow.

**User interaction control:** The achieved user interacts with the monitoring mechanism during runtime enable users to change program behaviors or modify the way that information flow while the program is executing. To our knowledge these have not been done before.

## II. RELATED WORK

Static verification involves the analysis of a source text by humans or software which can help to discover errors in the earlier stage of the software process [4]. Non-functional requirements (Security requirements) in information systems change more frequently than functional requirements especially when new users or new data are added to the system. Runtime verification [3, 5, 6] has been used to increase confidence that the system implementation is correct by ensuring that it conforms to its specification at runtime. Similar to [7] we employ runtime verification for the information flow to determine whether a flow in a given program run violates a security policy. The problem is to prevent sensitive information such as credit card data, government intelligence or personal medical information, from leaking through the computation into untrusted data sinks. The specifications of what information flows are considered dangerous are set out in the security policy which represents a formal description of the user's security

requirements. To manage change in security requirements and context of the security mechanisms our approach concentrates on providing a dynamic and adaptable information security solution by interacting with the user while the system is running in response to information flow events. Despite a long history and a large amount of research on information flow control [8, 9, 10, 11, 12] there seems to be very little research done on dynamic information flow analysis and enforcing information flow based policies. One of the fundamental work in this area is a lattice model of information flow which was presented in [13] by Denning. Considering this model Denning and Denning [9] provide static certification mechanism for verifying the secure flow of information through a program. Other approaches, eg. JFlow [14] and FlowCaml [12] use typing systems, checking information flow statically and work as source to source translators. Dynamic information flow analysis [15, 16, 17, 18] has been less developed than static analysis. Dynamic analysis began much earlier in the 1970s by Bell and LaPadula who aimed to deal with confidentiality of military information [19]. In their model they annotate each element of data with a security label and dynamically controlling information flow with two security properties. The simple security property 'no read up' which means any process can read from higher level security. The star property 'no writes down' which means that the process does not allow writing data to lower security level. Fenton [15] motivated research on dynamic analysis on a code level by his abstract data mark machine which dynamically checks information flow where each variable and program counter is tagged with a security label. Brown and Knight [16] provide a set of hardware mechanisms to ensure secure information flow. Lam and Chiueh [17] proposed a framework for dynamic taint analysis for C programs. In addition Birznies [18] provides an execution mode called Perl Taint Mode which is a special mode in the Perl script language where data are tagged with taint or untainted security level in order to detect and prevent the execution of bad commands. Vachharajani. et.al [20] proposed a framework for user centric information flow security at a binary code level; in this mechanism every storage location is associated to a security level. Cavadini and Cheda [10] presented two information flow monitoring techniques that use dynamic dependence graphs to track information flow during run-time. Previous work did not take in to consideration the ability for users to modify the flow policy at runtime and the security requirements to be dependent on the requirement of individual users and their interaction with the monitoring system. Many tools use techniques based on program instrumentation to carry out different tasks such as program tracing and optimization [26, 27]. Many tools developed using program instrumentation techniques that used for studying program behavior such as Pixie [28], Epoxie [29] and QPT [30] that rewriting program executables to generate address traces and instruction counts. MPTRACE [31] and ATUM [32] techniques used to report information about the timing and structure of the program. Also, there are software testing and

quality assurance tools that detect memory leaks and access errors such as Purify [33] catch programming errors by using these techniques. Purify inserts instructions directly into the object code produced by existing compilers. These instructions check every memory read, write performed by the program, to report any error occurs. One of the limitations of these tools that are designed to perform a specific task and are difficult to modify to meet users needs. Modification of a customized tool to obtain less of more trace information requires a user to access tool source code and understand the low level details of the tool. The above mentioned tools operate on object codes for a variety of operating systems and architectures, but none of them works on JVM class files. There is a tool called NetProf [34] that visualizes Java profile information by translating Java byte codes to Java source code. BIT [35] is a set of interfaces that brings the functionality of ATOM [36] to the Java world by allowing a user to instrument a JVM class file to observe the runtime behavior of programs for optimization. Furthermore, these are also customized tools and did not provide any trace information about the information flow that may happened within the application. In fact, most instrumentation framework [34, 37, 38] void instrumenting system classes, modifying only user created classes or limit their functionality to only user classes and FERRARI [39] instruments all system classes statically including those that are never used by an application.

### III. INFORMATION FLOW ANALYSIS VS. RUNTIME MONITORING

Static information flow analysis verifies whether programs can leak secret information or not, without running the program. Static analysis checks all possible execution paths including rarely executed ones. The advantage is that any program that has been successfully checked is indeed safe to execute as it cannot possibly leak information. The disadvantage is that any change in the underlying information flow policy means that the whole program needs to be analysed again. Another disadvantage is that a given program may be capable of leaking information, but in the way that it is executed by the user such leaks do not occur. Using static verification this program would be regarded unsafe. Dynamic information flow analysis is concerned with monitoring and regulating a program execution at run time. It is potentially more precise than static analysis because it only requires that the current execution path does not leak information and can also handle language features such as pointers, arrays and exceptions easier than static analysis. Finally, using runtime monitoring of information flow it is possible to allow for user interaction that can influence the further execution of the program.

Graphically we can depict the set of all possible program behavior by a blank circle and the set of all insecure program behaviour (defined in the policy) by the shaded circle.

In these terms a program is rejected by static analysis if the intersection of both is not empty. In Figure 1 we depict the case for dynamic information flow analysis. Consider that a

program is in a state 0 and performs an operation  $\alpha$  that causes an information flow. We can distinguish two cases:

1. After executing  $\alpha$  program is in a secure state.
2. After executing  $\alpha$  program is in an insecure state.

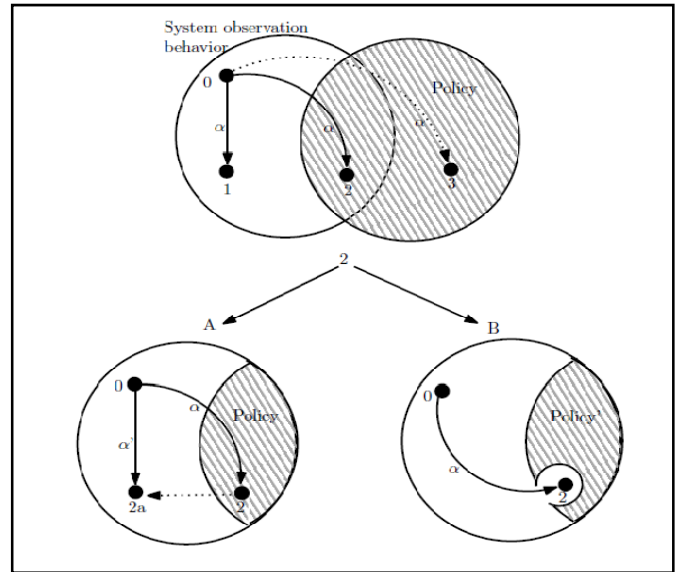


Figure 1. Runtime Monitoring

The hypothetical third case that the program exhibits a behavior that is defined by the policy as insecure, but is outside of the set of possible behaviours and so can be ignored. In our framework Figure 2 we check whether the program is about to enter an insecure state by intercepting the operation  $\alpha$ . In case 1, in which  $\alpha$  leads to another secure state the program will be allowed to perform  $\alpha$ . In case 2, the runtime monitoring mechanism will send feedback to the user asking about the violation of information flow policy. The user has two options on how to proceed:

(A) s/he changes the operation  $\alpha$  to another operation  $\alpha'$  in such a way that the resulting state is secure with respect to the policy. Such changes can for example be the termination of the program or (manually) sanitizing the information that flows in  $\alpha$ .

The other option (B) is to modify the Policy into a Policy' for which  $\alpha$  leads to a secure state. This could, for example, be introducing a one-of exception for the current flow or defining a separate context in which the information flow is considered legal.

### IV. FRAMEWORK

Our framework Figure 2 is designed to address government and military security needs; today commercial operations such as educational institutions, network service providers, companies, banks, and others are increasingly interested in enhancing security of their confidential information. Suppose that a program (attacker) requires a piece of confidential information from servers on the internal domains; Can we make sure that the information is not somehow being leaked?

Simply trusting the program is dangerous as we cannot always trust its provider. A better approach is to execute the program in a safe environment and monitor its behaviour to prevent confidential information from flowing to untrusted entities. However, the limitation is that the flow can only be controlled within the monitored program. Our approach will detect this violation of information flow and ask the user how to proceed. Our approach is based on the observation of information flow during application run time. For example given a policy that states that user Bob's password must not be shared with any other user (Alice and Eve). If Bob now wants to give his password to Alice to perform some function on his behalf, our approach will detect this violation of information flow and ask the user how to proceed. Bob can then choose to stop the operation that would violate the flow (i.e. Alice does not obtain the password) or he allows this flow and changes the policy accordingly to reflect this decision. Our framework involves several components as illustrated in Figure 2.

#### A. Security Requirements Specification

Stakeholders normally have a number of concerns that come with their desired system and are typically high-level strategic goals. In this component the stakeholders specifies the desired characteristics that a system or subsystem must possess with respect to sensitive information flow. In our previous example, this is the requirement that Bob's password must not be shared with any other user (Alice and Eve). In this stage the stakeholders should provide the source of the information and the output channel or a sink which will be formally expressed in the information flow policy.

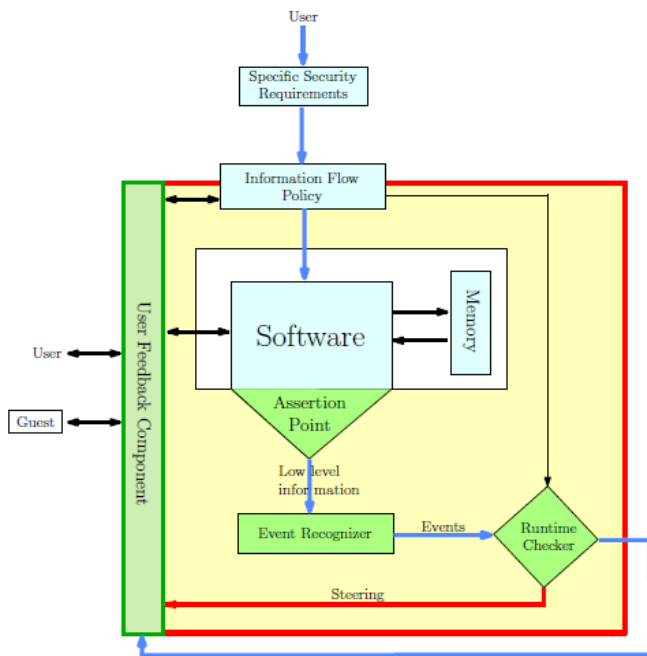


Figure 2. Runtime verification of information flow.

#### B. Information Flow Policy

Information flow policy is a type of security policy that defines the authorized paths along which information can move throughout the system, a policy is set of rules, axioms and laws that regulate how information must flow to prevent a leak of sensitive information. The information flow policy is designed to preserve information confidentiality or integrity. In our framework, the information flow policy expresses the security requirements as specified by the stakeholder/user to a set of rules that are understandable by our runtime monitoring mechanism. An information flow policy describes the events at the requirement level. The goal of information flow policy is to prevent secret information from flowing to a user or client software not authorized to receive it. e.g the information flow policy can state more context dependent flows, such as "Bob's password can only flow to Alice if Bob explicitly agrees".

#### C. Assertion Points

Assertion points are program fragments as a collection of probes that will be added into the target software.

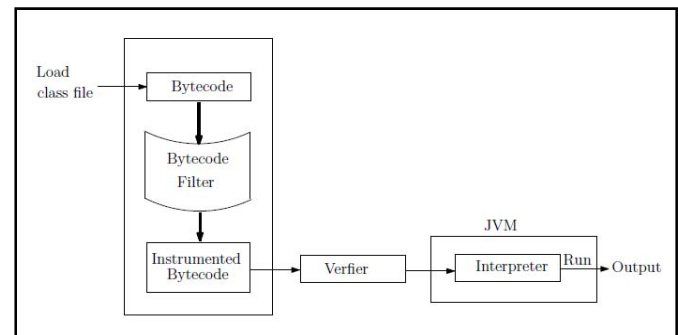


Figure 3. Assertion point process (instrumentation)

The main functionality of the assertion point is to send the relevant state information to the event recognizer. This will ensure the monitoring of relevant objects during the execution of the software. The probes are inserted into all locations where monitored objects are updated such as (program variables and function calls); unlike traditional runtime verification approaches, our assertion points are inserted before the operation to be able to intercept updates and thus prevent the system from entering an insecure state. Due to sending proper state information to the Event recognizer our framework uses a new instrumentation mechanism. The assertion point process is deployed using Javassist [25].

#### D. Event Recognizer

An event recognizer is the part of the framework that detects an event from the information received from the assertion point. The Event recognizer is used as a communication interface between the assertion points and the runtime checker. The event recognizer sends any event that attempts to change the state of the software (according to the information flow policy). In addition to sending events, the

event recognizer can also send the value of the variables to the runtime checker to use them for checking against information flow policy and to provide intelligible feedback to the user. Although it is possible to combine these two components the assertion point and the event recognizer, we separated them to make the implementation of our architecture more extensible and to minimize the interference with the monitored application, such as allowing the integration of several runtime checkers that verify different type of requirements. For example, the management of obligations related to information flow could be placed in a logically separate component

#### E. Runtime Checker

The runtime checker checks that the execution trace belongs to the set of all acceptable behaviours as defined by the security requirements specification and expressed in the information flow policy. It verifies and decides whether or not the current execution trace as obtained from the event recognizer satisfies the information flow policy and sends feedback to the user feedback component when it determines that the software is about to enter an insecure state e.g. any information flow that violate the policy.

#### F. User Feedback Component

The user feedback component is an interface between our system and the user. The user feedback component handles all interactions with the system and the user. It runs in a separate thread of control so that user interaction can be overlapped with information flow control. The user feedback component also allows the user to administrate the policy. When the software is running, the user feedback component receives feedback from the runtime checker (Steering). The user feedback component informs the user about any feedback received from the runtime checker. As illustrated in Figure 2 if the runtime checker determines that this state execution would violate the information flow policy then it sends feedback to the user, the system behavior will be changed accordingly, and the policy will be modified according to the user decision.

#### G. Java Agent

The Java agent is deployed as a JAR or ZIP file. An attribute in the agent file manifest specifies the agent target class which must be loaded to start the agent. In our framework the Java agent is started by specifying an option on the command line. Implementations may also support a mechanism to start agents some time after the JVM has been started. Figure 4 shows the command line interface for starting our runtime mechanism.

```
$ java -javaagent:JavaAgent.jar uk.ac.dmu.msarab.vif.testtargets.HelloWorld
```

Figure 4. Command line interface

### V. RUNTIME MONITORING MECHANISM

To achieve our approach objectives two main steps are taken:

- Loading and instrumentation of target program class file.

- Executing the main class file and monitoring information flow with respect to the information flow policy.

Figure 5 shows a flowchart of our runtime monitoring mechanism steps.

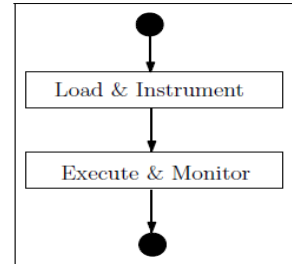


Figure 5. Monitoring mechanism flowchart

#### A. Loading and Instrumentation

Loading is the first operation performed by class loader. The aim of loading process is to obtain a binary form of a class file [21, 22, 23]. The main components of this process are Java Class file and Class loader. The Bytecode instrumentation is process that inserts method calls to analysis the bytecode, such that information can be extracted from the target program while it is being monitored and executed. Our approach uses dynamic instrumentation mechanism whereas all classes that actually loaded will be instrumented including core classes of JDK. Class loader can only provide the classes it defines, not the classes delegated to other classes loader. Thus our approach uses Java agent to allow all loaded classes (bytecode) to be instrumented and redefined on Java virtual machine during runtime. Our instrumentation mechanism is general enough to be applicable to any Java bytecode and is flexible mechanism to send potential state information to the event recognizer.

#### B. Executing and Monitoring

The second step of our runtime monitoring mechanism is to execute the class file and monitor the flow of the information flow with respect to the information flow policy. In our runtime monitoring mechanism a new runtime frame is created each time a method is invoked. The runtime frame involves a new stack called information flow stack (IFS) and Symbol Table for the use by the current method to store its variables. At any point of the execution, there are thus likely to be many frames and equally many information flow stacks (IFS) and Symbol tables per method of control, corresponding to many nested method invocations. Only the runtime frame (IFS and Symbol table) of the current method is active. The event recognizer receives an event that attempts to change the state of the information flow within the application. The event recognizer manipulates all variables and their values using the current runtime frame (IFS and Symbol table) and implicit information flow stack (IMFS) for controlling implicit information flow. The implicit information flow stack (IMFS)



is sharing between all runtime frames as illustrated in Figure 6.

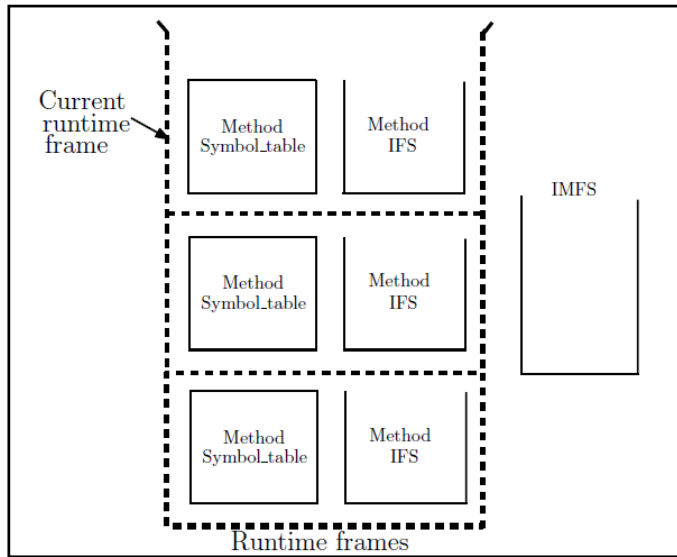


Figure 6. Runtime frame of the current method and IMFS

The Framework prototype implementation can be found in: <http://www.tech.dmu.ac.uk/~msarrab>

## VI. CASE STUDY

The presented case study is hello world Java program to demonstrate certain aspects of our information flow control approach in more details. The case study has been chosen to be a small to show how the Instrumentation process, Event recognizer and Runtime checker components interact together to figure out any possible information flow within java application. The case study Figure 1 consists of one java class named Test.java

```
public class HelloWorld{
    public static void main(String[] args) {
        System.out.print(`Hello World`);
    }
}
```

Figure 7. *Source code of HelloWorld.java*

The given program Figure 7 contains only one method named *HelloWorld.main()*. The class execution starts at the main method first line. Statements are executed one at a time, as ordered in the main method, until reached the end of the method or another method invocation. As shown in line 3 in our example Figure 7 *java.io.PrintStream.print()* method. Method invocations are like a detour of the execution flow. Figure 8 shows all possible execution flow of class *HelloWorld.java*. Instead of perform the next statement 4 in the current method *HelloWorld.main*, the execution flow to the first statement in the invoked method *java.io.PrintStream.print()*, executes all the statements in the invoked method, and then back and start again where the mother method left off with new statement which is in our example Figure 7 line 4. That sounds simple enough, except that a method can invoke another method.

```
Test.main()
→ java.io.PrintStream.print()
    → java.io.PrintStream.write()
        → java.io.PrintStream.ensureOpen()
            → java.io.Writer.write()
                → java.lang.String.length()
                    → java.io.Writer.write()
                        → java.lang.String.getChars()
                            → java.lang.System.arraycopy()
                                → java.io.Writer.write()
→ java.io.BufferedWriter.flushBuffer()
    → java.io.BufferedWriter.ensureOpen()
        → java.io.Writer.write()\\
→ java.io.OutputStreamWriter.flushBuffer()
    → sun.nio.cs.StreamEncoder.flushBuffer()
        → sun.nio.cs.StreamEncoder.isOpen()
            → sun.nio.cs.StreamEncoder.implFlushBuffer()
                → sun.nio.cs.StreamEncoder.writeBytes()
                    → java.nio.channels.WritableByteChannel.write()
                        → java.io.OutputStream.write()
                            → java.io.OutputStream.write()
                                → java.nio.Buffer.clear()
                                    → java.nio.Buffer.clearMark()
→ java.lang.String.indexOf()
    → java.lang.String.indexOf()
→ java.io.OutputStream.flush()
```

Figure 10. *The instrumented bytecode of the HelloWorld.java*

After instrument all loaded classes where assertion points are inserted in the bytecode. The instrumented classes are ready to execute. As mentioned in Executing and monitoring 5.2 when *HelloWorld.main()* start execute our run time mechanism creates a new implicit flow stack (IMFS) and a new runtime frame Figure 11.

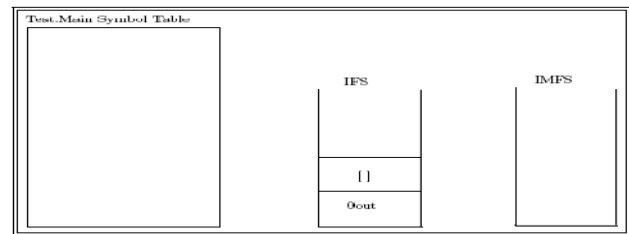


Figure 11. *Runtime frame and IMFS of the HelloWorld.Main()*

As depicted in Figure 10 line 02 sends an event to the Event recognizer to load field named *0out*. The Event recognizer pushes *0out* onto the IFS. The second event is in line 08 that sends to load constant. The Event recognizer pushes empty string onto the top of the IFS. The third event is at line 17 that sends to inform the Event recognizer that another method is about to invoke named *print*, 1 parameter and 0 return value as specified in lines 13,15 and 16 respectively. The Event recognizer will create new runtime frame for the *print* method. The *print* method symbol table will contain the object reference and the parameter in location 0 and 1 because the Event recognizer will pop the top two elements from the current method *HelloWorld.Main()* IFS and checks the IMFS if it has any element to combine with each popped element from the IFS to handle the implicit flow. In our example the IMFS is empty. Then the Event recognizer sends them to the *Print* method symbol table as illustrated in Figure 12.

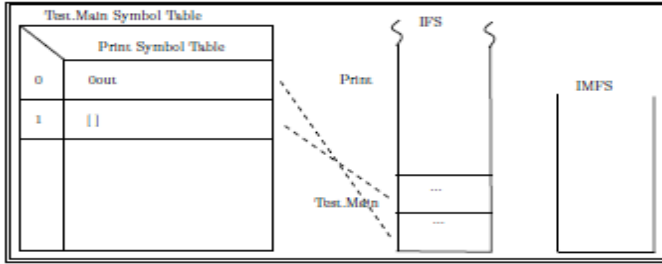


Figure 12. Runtime frame and IMFS of the `java.io.PrintStream.print()`

```

00: aload_1
01: dup
02: ldc_w #1249 = "1"
05: invokestatic #1251 = Method Monitor.Event Recognizer.
    Aload((Ljava/lang/Object;Ljava/lang/String;)V)
08: invokestatic #1253 = Method Monitor.EventRecognizer.
    ifcond((V)
11: ifnonnull 30
14: invokestatic #1255 = Method Monitor.EventRecognizer.
    Const((V)
17: ldc #52 = "null"
19: dup
20: ldc_w #1256 = "1"
23: invokestatic #1258 = Method Monitor.EventRecognizer.
    Store((Ljava/lang/Object;Ljava/lang/String;)V)
26: astore_1
27: invokestatic #1270 = Method Monitor.EventRecognizer.
    Endif((V)
30: aload_0
31: dup
32: ldc_w #1259 = "0"
35: invokestatic #1261 = Method Monitor.EventRecognizer.
    Aload((Ljava/lang/Object;Ljava/lang/String;)V)
38: aload_1
39: dup
40: ldc_w #1262 = "1"
43: invokestatic #1264 = Method Monitor.EventRecognizer.
    Aload((Ljava/lang/Object;Ljava/lang/String;)V)
46: ldc #1244 = "write"
48: iconst_1
49: iconst_0
52: invokestatic #1266 = Method Monitor.EventRecognizer.
    Method((II)V)
55: invokespecial #45 = Method java.io.PrintStream.
    write((Ljava/lang/String;)V)
58: iconst_1
59: invokestatic #1268 = Method Monitor.EventRecognizer.
    Return((I)V)
62: return

```

Figure 13. The instrumented bytecode of `java.io.PrintStream.print()`.

As indicated in Figure 13, line 05 sends an event to the Event recognizer to load contains of Symbol table location 1 as specified in line 02.

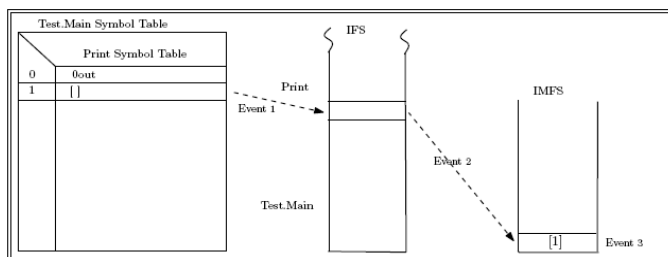


Figure 14. Snapshot of the runtime frame and IMFS of current method `java.io.Print`.

The Event recognizer pushes label {1} onto the current runtime frame IFS. The second event is at line 08 that sends to inform the Event recognizer about if statement. The Event recognizer pops one element from the top of IFS which is in our case label {1} and pushes it onto the top of the IMFS as illustrated in Figure 14.

The next event will be sent according to the condition at line 11. Assuming that the condition is false where contains of local variable location 1 is not null; the execution flow will jump to line 27. Because or mechanism will reduce if statement offset address by 3 to execute Endif. The third event is at line 27 that sends to inform the Event recognizer that if statement ends at line 27, then the Event recognizer pops the top element on the IMFS. Now both stacks (IFS and IMFS) are empty. The next event is at line 35 that sends to load contains of Symbol table location 0 as specified in line 32. The Event recognizer pushes label {0} onto the IFS. The next call to the Event recognizer is at line 43 to load contains of Symbol table location 1 as specified in line 40. Again the Event recognizer pushes label {1} onto the IFS. The next event is at line 52 that sends to inform the Event recognizer that another method is about to invoke named `write`, 1 parameter and 0 return value as specified in lines 46, 48 and 49 respectively. The Event recognizer will create new runtime frame for the `write` method. The Event recognizer will pop the top two elements ({1}, {0}) from the current method IFS which are labels to the contains of the current method Symbol table and checks the IMFS if there is any element to combine with the popped elements from IFS, in our case IMFS has no element. Then the Event recognizer sends their values to the `write` method Symbol table as depicted in Figure 15. The execution will continue in this way creating runtime frame for each invoked method and pass the parameters between the methods frames until return from the method to destroy its runtime frame, where our runtime monitoring mechanism will intercept.

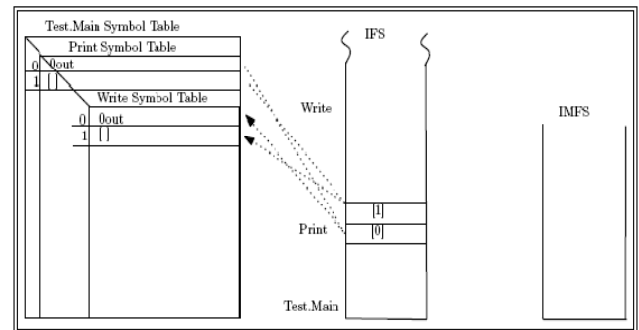


Figure 15. Snapshot of the runtime frame and IMFS of current method `java.io.Writer.write()`

Line 232 Figure 16 sends an event to the Event recognizer to load label 0 as specified in line 229. The Event recognizer pushes label {0} onto the top of IFS of the current method `Write`.

```
227: aload_0
228: dup
229: ldc_w #336 = "0"
232: invokestatic #338 = Method Monitor.EventRecognizer.
    Aload((Ljava/lang/Object;Ljava/lang/String;)V)
235: aload 5
237: dup
238: ldc_w #339 = "5"
241: invokestatic #341 = Method Monitor.EventRecognizer.
    Aload((Ljava/lang/Object;Ljava/lang/String;)V)
244: invokestatic #343 = Method Monitor.EventRecognizer.
    Const(0)V
247: iconst_0
248: iload_3
249: dup
250: ldc_w #344 = "3"
253: invokestatic #346 = Method Monitor.EventRecognizer.
    Iload((Ljava/lang/String;)V)
256: ldc_w #347 = "write"
259: iconst_3
260: iconst_0
261: invokestatic #349 = Method Monitor.EventRecognizer.
    NativeWrite((Ljava/lang/String;II)V)
264: invokevirtual #7=Method java.io.Writer.write ((CII)V)
```

Figure 16. A snapshot of instrumented bytecode of *java.io.Writer.write*

Line 241 sends another event to load label 5 as specified in line 238. The Event recognizer pushes label {5} onto the top of IFS. Line 244 sends an event to load constant. The Event recognizer pushes empty string {} onto the IFS. Next event at line 253 sends to inform about loading label 3. The Event recognizer pushes label {3} onto the top of IFS as shown in Figure 9.5. Line 261 sends to inform the Event recognizer that **NativeWrite** method is about to invoke named *java.io.Write.write*, 3 parameter and 0 return value as specified in lines 256, 259 and 260 respectively.

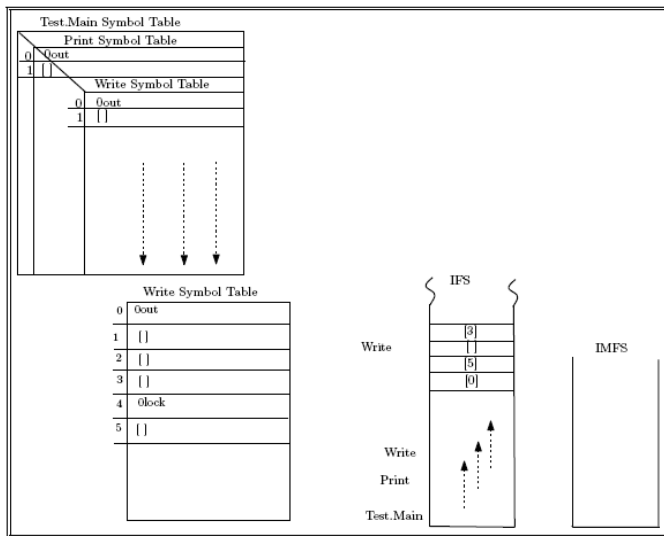


Figure 17. A snapshot of the current method run time frame at line 261

The Event recognizer sends the contains of the labels {0}, {5}, {} and {3} (Out, , , ) to the Runtime checker and the Runtime checker will figure out that empty string is going to flow to the System.out. In this case no need to check the information flow policy and send a message to the User Feedback component. Therefore, the execution will continue as normal without any

intercept from our run time mechanism. Figure 17 shows the run time frame contains of the current method *java.io.Writer.write* when the execution at line 261. As result all our framework components were operated exactly as intended. However, the limitation of our approach is that the user will be altered about the way that information may flow in but nothing about the amount of the flow is provided.

## VII. CONCLUSIONS

This article presents policy-based sandbox architecture for monitoring information flow at the application level. The benefits of this approach are two-fold. Firstly, monitoring information flow at runtime has the advantage over static verification methods such as [3, 5, 6] that it is possible to interact with a user and therefore allow for more flexible control to be exercised. Secondly, our approach does not treat the application as a black box (with the general assumption that once information has passed into it can find its way to any destination). Instead the actual flows that take place at runtime are traced and the program is only interrupted when a policy violation does occur. This means that even "unsafe" programs may be executed within "safe" parameters, i.e. as long as they do not actually violate the information flow policy. Static verification on the other hand would rule out these programs from the outset, as they can potentially violate the policy. Our approach provides a higher degree of flexibility to support:

- Users' ability to modify the information flow policy during runtime.
- Detecting and monitoring of potential leaking behaviour of a program and the user decides whether to abort or continue the program execution.

Our approach ensures that the program contains only those flows approved by the user.

## REFERENCES

- [1] <http://docs.sun.com/app/docs/doc/8058115-10/6j7klujb8>.
- [2] R. Anderson. Security Engineering, A Guide to Building Dependable Distributed Systems, Second edition, UK, 2001.
- [3] H. Janicke, F. Sieve, K. Jones, A. Cau and H. Zedan. Proceedings of the Workshop on Defence Applications & Multi-Agent Systems (DAMAS05), in 4th international joint conference on Autonomous Agents & Multi Agent Systems (AAMAS05), July 25, 2005.
- [4] K. Havelind and A. Goldberg, "Verify your runs", Verified Software: Theories, Tools, Experiments, 2005.
- [5] H. Ben-abdallah S. Kannan L. Insup O. Sokolsky M. Kim, M. Viswanathan, "Mac: A framework for run-time correctness assurance of real-time systems", In Philadelphia, PA, Department of computer and Information Science University of Pennsylvania., 1999.
- [6] S. Kannan M. Kim O. Sokolsky M. Viswanathan L. Insup, H. Ben-abdallah. "A monitoring and checking framework for run-time correctness assurance" 1998.
- [7] S. Cavadini and D. Cheda. "Run-time information flow Monitoring based on dynamic dependence graph", In IEEE Computer society, Third International Conference on Availability, Reliability and Security, 2008.
- [8] A. Banerjee and D. A. Naumann. "History-based access control and secure information flow", Lecture Notes in Computer Science, 2005, Vol.3326, pp. 27-48.

- [9] D. E. Denning and P. J. Denning, "Certification of programs for secure information flow", Commun. ACM, 1977, Vol.20, No.7, pp. 504-513.
- [10] D. Volpano, G. Smith, and C. Irvine, "A sound type system for secure flow analysis," J. Computer. Security, Vol. 4, No. 3, 1996, pp. 167-187.
- [11] G. Smith and D. Volpano. "Secure information flow in a multi-threaded imperative language", Proceedings of the 25th ACM SIGPLAN-SIGACT, 1998, pp. 355-364.
- [12] F. Pottier and V. Simonet, "Information flow inference for ML", in Proc. ACM Symp. Principles Programming Languages, 2002, pp. 319-330.
- [13] D. E. Denning, "A lattice model of secure information flow," Commun. ACM, Vol. 19, No. 5, 1976, pp. 236-243.
- [14] C. Myers, "JFlow: Practical mostly-static information flow control" In proceeding of 26th ACM Symposium on Principles of Programming Language, 1999, pp. 228-241.
- [15] J. Fenton, "Memory less subsystems", The Computer Journal, Vol.17, No. 2, 1974, pp. 143-147.
- [16] J. Brown and JR. F. Knight. A minimal trusted computing base for dynamically ensuring secure information flow. Technical Report ARIES-TM-015, MIT, 2001.
- [17] L. Chung and T. Chiueh, "A general dynamic information flow tracking framework for security applications", Washington, DC, USA. 2006, pp 463-472.
- [18] G. Birnieks. "Perl taint mode version 1.0" june 3, 1998 <http://gunther.web66.com/faqs/taintmode.html>. 1988.
- [19] D. E. Bell and L. J. LaPadula, "Secure computer systems: Mathematical foundations", Tech. Rep. MTR-2547, Vol. 1, 1973 MITRE Corp., Bedford.
- [20] N. Vachharajani, M. J. Bridges, J. Chang, R. Rangan G. Ottoni, J. A. Blome, G. A. Reis, M. Vachharajani, and D. I. August. Rifle: "An architectural framework for user-centric information-flow security", In Proceedings of the 37<sup>th</sup> International Symposium on Micro architecture, 2004.
- [21] T. Dowling and J. Meyer. Java virtual machine. O'Reilly and Associates, Inc. Sebastopol, CA, USA, 1997.
- [22] F. Yellin, T. Lindholm, The JavaTM Virtual Machine Specification. Addison-Wesley Longman Publishing Co, Inc, 1997.
- [23] F. Yellin T. Lindholm, The JavaTM Virtual Machine Specification, Addison-Wesley Longman Publishing Co, Inc., 1999.
- [24] M. Bishop, Computer Security Art and Science, Addison-Wesley, 2002.
- [25] S. Chiba, "Load-time structural reflection in Java", Lecture Notes in Computer Science, Vol.1850, 2000, pp.313-336.
- [26] A. Srivastava and D.Wall, "Link-time optimization of address calculation on a 64-bit architecture", In Proceedings of the SIGPLAN 94 Conference on Programming Language Design and Implementation (PLDI), 1994, pp 49-60.
- [27] A. Srivastava and D.Wall, "A practical system for intermodal code optimization at link-time". Journal of Programming Languages, Vol.1, No.1, 1993, pp 1-16.
- [28] M. Smith, Tracing with pixie. memo from center for integrated systems, Stanford university. April 1991.
- [29] D. Wall, "Systems for late code modification tools techniques", Springer-Verlag, 1992, pp. 275-293.
- [30] R. Larus and T. Ball. "Rewriting executable files to measure program behavior", Software Practice and Experience, Vol.24, No.2, pp.197-218.
- [31] B. Peter S.Kessler L.Graham and K. McKusick. "An execution profiler for modular programs ", Software Practice and Experience, Vol.1, No.13, 1993, pp.671-685.
- [32] L. Sites A.Agarwal and M.Horowitz, "Atum: A new technique for capturing address traces using microcode". Proceedings of the 13th International Symposium on Computer Architecture, 1986, pp 119-127.
- [33] R. Hastings and B. Joyce, "Purify: Fast detection of memory leaks and access errors", Proceedings of the Winter USENIX Conference, 1992, pp125-136.
- [34] M. Cierniak S.Parthasarathy and L. Wei, "Netprof: Network based high-level profiling of java bytecode", Technical Report 622, Computer Science Department, University of Rochester, 1996.
- [35] L. Bok and G. Benjamin, "Bit:A tool for instrumenting java byte codes", Proceedings of the USENIX Symposium on Internet Technologies and Systems Monterey, 1997.
- [36] A. Srivastava and A. Eustace, "Atom a system for building customized program analysis tools", In Proceedings of the SIGPLAN 94 Conference on Programming Language Design and Implementation (PLDI), 1994, pp. 196-205.
- [37] S. Chiba M. Tsubori T.Sasaki and K.Itano, "A bytecode translator for distributed execution of legacy". Lecture Notes in Computer Science, 2001, pp.20-70.
- [38] E. Tilevich and Y.Smaragdakis. "J-orchestra: Automatic java application partitioning". In European Conference on Object-Oriented Programming (ECOOP), Spain, 2002.
- [39] P. Moret W. Binder and J. Hulaas, "Reengineering standard java runtime systems through dynamic bytecode instrumentation", Proceedings in Seventh IEEE International Working Conference on Source Code Analysis and Manipulation. Computer Society Washington, DC, USA, 2007, pp. 91-100.

#### AUTHORS PROFILE



Computer forensic.



Quality Assurance

**Mohamed Sarrah** is a PhD student in De Montfort University. Faculty of Technology. Software Technology Research Laboratory (STRL). He received his B.Sc in computer science from 7<sup>th</sup> April University Libya and also received M.Sc in Computer Science from VSB Technical University of Ostrava CZ. His main research interests are Computer security, Runtime Verification,

**Helge Janicke** received his Ph.D. degree in Computer Security from De Montfort University 2007. He is working as a Senior Lecturer in Computer Security at De Montfort University. Faculty of Technology. Software Technology Research Laboratory (STRL) His main research interests are in Computer Security, Policy Based Management, Formal Methods and Software

# Design of Novel Usability Driven Software Process Model

Hina Gull

Department of Computer Engineering,  
College of Electrical & Mechanical Engineering  
NUSTRawalpindi Pakistan  
hina.gull03@gmail.com

Farooque Azam

Department of Computer Engineering,  
College of Electrical & Mechanical Engineering  
NUSTRawalpindi Pakistan  
farooq@ceme.nust.edu.pk

Sardar Zafar Iqbal

Department of Computer Sciences,  
Iqra University, Islamabad Pakistan  
zafar.iqbal38@gmail.com

**Abstract-** In this paper we have proposed a novel software process model for web based applications. This model is based on the empirical study carried out by us and also by the literature review of software development models. Model consists of three phases: Requirement Engineering, Design and Implementation. Model contains certain sub activities in each phase describing the flow and steps which should be followed to develop a web application. The main emphasis of the model is on usability aspect, keeping in view the criticality of the user interface for a good web application. Flexible and easy change manageable nature of the model makes it different and worth using as compared to other software development approaches.

## I. INTRODUCTION

The rigorous use of web application has produced a mounting interest in the development of methodological approaches providing a suitable support for the construction of web applications and other software applications. Several research groups proposed methodologies, processes and models to build software applications. Several software process models (discussed in Section II) were studied and analyzed properly and thoroughly. The analysis revealed that there are some limitations in the existing models. Hence there arises a need to development of new customized model to accommodate the needs of any web application.

This paper presents novel usability driven web based software process model, which proposes to build the web application in three phases, containing back and forth movement from one phase to another to accommodate changes at any phase. The paper contains the core methodology behind the model and different phases, how they relate to each other and how they contribute to web application design and structure. The paper is organized as follows: Section II contains the background study of existing

software process models. Section III contains the proposed model Section IV contains the Conclusion and Future Work.

## I. BACKGROUND

A set of activities whose goal is the development and evaluation of a software product is called a software process. [1]. General definition given by the commonly accepted Software Engineering Institute's Capability Maturity Model (SEI CMM) is "A software process is a set of activities, methods, practices, and transformations used by people to develop software"[2]. Software development process constitutes the following activities:

- Requirements engineering: aims at understanding the problem.
- Design: aims at planning a solution to the problem.
- Implementation: translates the plan into running application code.
- Testing: aims at identifying coding errors or inconsistencies between the collected requirements and their implementation.
- Deployment: brings the solution to the customers.
- Maintenance: aims at monitoring a running system and keeping it healthy and running.

A literature review of history of software development life cycle dates back to 1960s. Working for the US department of defence A. Enthoven and Henry Rowan developed a heuristic process for managing large information system project [3]. This process defines linear set of stages for project development that could be recycled as needed. Winston Royce introduced the first model for software development in 1970 in his paper "Managing the development of large software systems" [4]. Following is the description of the model given by him:



#### A. Waterfall Process Model

Waterfall model is recognized as Classic Life Cycle Model or Linear Sequential Model. This was the first stepwise sequential model given in 1970 by Winston W. Royce. Because of the cascading effect from one phase to another this model is known as the waterfall model. The model formed the basis for most software development standards and consists of the following phases: Requirements specification, Design, Coding, Testing and Debugging, Installation and Maintenance.

#### B. V Software Process Model

V Model is considered to be the extension of the waterfall model. After moving in the linear fashion the process steps are bent upward after the coding phase. For top-down SE (i.e., forward engineering), the process starts on the upper left and goes to the upper right. For bottom-up (i.e., reverse engineering), it starts on the upper right and goes to the upper left. It is commonly used to identify the relationship between the development phase and the testing phase. V- Model is shown in figure given below. The model follows a well structured method in which each phase is implemented by the detailed documentation of the previous phase [5].

V- Model consists of number of phases. At the left side of the model, verifications phases are there, coding is at the bottom and validation phases are at the right side of the V.

#### C. Incremental Development Model

“The incremental model delivers software in small, usable pieces, called “increments. In general, each increment builds on those that have already been delivered.” (Roger S. Pressman).

Waterfall model of the software development requires its users to commit the requirement phase before the design begins and the designer to commit specific design strategies to commit design before implementation. Changes to the requirements require rework on requirements, design and implementation [6]. Keeping these weaknesses of the waterfall in view incremental development model was developed. It sets off with an initial planning and ends with deployment with the cyclic connections in between.

As calendar time progresses, Incremental model applies linear sequences in a staggered fashion. Each linear sequence delivers an increment of the software. All requirements are determined initially and individual increments are allocated with subsets of requirements. The increments are developed in sequential series, with each incremental release adding functionality. The incremental approach lessens overall effort leading to the earlier delivery of initial system to the customer. Typically, there is more effort in requirements analysis and design, and less for coding and integration due to which the overall schedule may lengthen.

#### D. Prototype Model

The goal of a prototyping-based development process is to overcome the limitations of the waterfall model. The

basic idea was producing a throwaway prototype to help in understanding the requirements instead of freezing the requirements before any design or coding can proceed. The prototype is produced on the basis of currently available set of requirements. The prototype helps customer to get the real look and feel of the system, enabling them to better understand what they actually want, leading more required features of the system and less frequent changes [7]. Figure 2.7 given below shows the prototyping paradigm [8]. The paradigm begins with requirement gathering. Overall objectives of the system are defined in the meeting of developers and customers, identifying the main requirements for the system leading to a quick design. This quick design represents the aspects of the system that are visible to the users. The design leads to a prototype of the system. The prototype is given to customer/user for evaluation. After exploring the prototype customer/user provides developers with their feedback i.e. what is correct, what needs to be modified, what is missing, what is not needed, etc. Based on the feedback, the prototype is modified to incorporate some of the suggested changes that can be done easily, and then the users and the clients are again allowed to use the system. Iteration occurs as the prototype is tuned to satisfy the needs of the customer, while at the same time enabling the developer to better understand what needs to be done.

#### E. Spiral Model

Spiral Model [9] is an evolutionary development model, originally proposed by Boehm. It is said to be the risk driven model combining iterative nature of prototype model and systematic aspects of linear sequential model. It can accommodate other models as special cases and provides guidance to determine which combination of models best fits a given situation. The spiral model has evolved through practical experience after implementation of complex government projects with new innovations in waterfall model in several years. The radial dimension model represents the growing cost required in bring about the steps to date; the angular dimension track the evolution made in completing each cycle of the spiral. It is called risk-driven because it identifies areas of uncertainties that are sources of project risk and structures activities based on the risks. The development proceeds in repeating cycles of determining objectives, evaluating alternatives, prototyping and developing, and then planning the next cycle. Each cycle involves a progression that addresses the same sequence of steps for each portion of the product and for each level of elaboration. Development builds on top of the results of previous spirals.

Each cycle of the spiral commence with the recognition of

- The intention i.e. (performance, functionality, ability to accommodate change, etc.);
- The unconventional resources i.e. (design A, design B, reuse, buy, etc.);
- The constriction i.e. (cost, schedule, interface, etc.).

#### F. Extreme Programming (XP)

XP is considered to be the lightweight methodology proven to be the efficient, low-risk, flexible, predictable and scientific [10]. It is designed for smaller projects and teams of two to ten programmers, which results in efficient testing and running of given solutions in a small amount of time. Its incremental planning approach and solid and continuing feedback from short cycles of the software development allow an evolutionary design process that lasts as long as its system. XP believes that changes are expected and rather than treating changes as undesirable, development should squeeze change. And to accommodate change, the development process has to be lightweight and quick to respond. For this, it develops software iteratively, and avoids reliance on detailed and multiple documents which are hard to maintain. Instead it relies on face-to-face communication, simplicity, and feedback to ensure that the desired changes are quickly and correctly reflected in the programs [7].

In XP, the process starts with the requirements which are expressed as user stories which are short (a few sentences) descriptions of what scenarios the customers and users would like the system to do. They are different from traditional requirements specification as the user stories do not contain detailed requirements which are to be exposed only when the story is to be implemented, therefore allowing the details to be decided as late as possible. Each story is written on a separate card, so they can be flexibly grouped. The authorized development team estimates how long it will take to implement a user story. The estimates are rough, generally stated in weeks. Using these estimates and the stories, release planning is done which defines which system release will contain which stories to be implemented in. The dates of these releases are also decided. XP encourages frequent and small releases. Programmers work in pairs and develop tests (acceptance tests) for each task before writing the code. All tests must be successfully executed when new code is integrated into the system [11]. Bugs found during the acceptance testing for iteration can form work items for the next iteration.

#### G. Win Win Spiral Model:

In Win Win Spiral Model [8], the developer and customer end up in negotiating various requirements based on functionality, performance, cost, time, etc. The best negotiation strives for a “win-win” result. The detailed risk analysis imposing many different constraints, objectives & alternatives consume a lot of time. But never are these risks specifically mentioned and vary project to project [9].

## II. PROPOSED MODEL

#### A. What is Usability Driven Web based Software Process Model?

Proposed Usability Driven Web based software process model is a software development model especially meant for web based software, containing several stages or phases from static information

gathering to the web based software's testing and deployment. It is complete in the sense it appears as an entire process to build a web based software application. It covers all the activities rendered for software development from requirement elicitation to design, coding and testing. Model contains certain sub activities in each phase describing the flow and steps which should be followed to develop a web application. The main emphasis of the model is on usability aspect, keeping in view the criticality of the user interface for a good web application. Flexible and easy change manageable nature of the model makes it different and worth using as compared to other development approaches of web based applications.

Model progresses in linear fashion if no change arrives, but the moment as these changes come across, model find its way back to the phase wherever this change needs to be accommodated. This back and forth movement of the model makes the difference in any environment to accommodate change.

#### a. Model Principles

Following are some of the principles which laid the foundation of the novel web development model:

- Flexibility & Change Management: Organizations and information are often changing so quickly that the information provided on web sites soon becomes out of date. If a web site has not been updated for several months, the confidence of the users in the provided information will probably not be very high. Hence change management is main focus of the model to cope with the changes.
- Attain Usability: Main concern of the web application is the providence of a user friendly interface. Model promotes methodology that will lead to a user friendly web application development.
- Evolve by continuously obtain feedback and improve: This principle promotes practices that allow the team to get early and continuous feedback from stakeholders about requirements and other issues leading to correct and in time delivery of the web application.

#### B. How Usability Driven Web Based Software Process Model is organized?

Model can be organized into two linked dimensions: method content and process content. The method content is where method elements are defined (namely roles, tasks or sub phases, and artifacts). The process content is where the elements of method content are applied in a predefined and chronological manner.

*b. Method Content:*

- **Roles**

The essential skills needed by team members to develop a web application using this model are represented by Roles:

- **Stakeholder** represents interest groups whose needs must be satisfied by the Project developed by the model. It is a role that may be played by anyone who is (or potentially will be) materially affected by the outcome of the project
- **System Analyst** represents customer and end-user concerns by gathering input from stakeholders to understand the problem to be solved and by capturing and setting priorities for requirements.
- **Designer** is responsible for designing the software architecture and other design concerns (Class diagram, Activity Diagram, Domain Model, Presentation Model, Navigational Model), which includes making the key technical decisions that constrain the overall design and implementation of the project.
- **Developer** is responsible for developing and implementing the system.
- **Tester** is responsible for the core activities of the test effort, such as identifying, defining, implementing, and conducting the necessary tests, as well as logging the outcomes of the testing and analyzing the results.
- **Project Manager** leads the planning of the project in collaboration with stakeholders and team, coordinates interactions with the stakeholders, and keeps the project team focused on meeting the project objectives.
- **Any Role** represents anyone on the team that can perform general tasks.

- **Disciplines (Model Phases)**

Following are the main phases of the model:

- Requirement Engineering
- Design
- Implementation

The model mainly consists of three phases: Requirement Engineering, Design and implementation. Each phase is further sub divided into sub phases or activities which form the basic flow of the model. The main feature of the model includes its flexibility i.e. the movement from any phase to another. The forward flow of the model suggests that one can move from requirement engineering phase to Design phase, then to the implementation phase. The model does not freeze any phase, one can easily move from design to requirement engineering phase and from implementation to requirement engineering and design.

The model can easily accommodate changes even later in the development cycle, due to its ability to move to any prior phase. The major issue with the ERP development was its changing and upcoming requirements (as customers were

officials of E&ME, who keep sending new requirement as well as can change any requirement at any time), so this model includes requirement change management process (Figure 4) keeping in view the above mentioned issue of requirements, so that requirement changeability or new requirements can be easily accommodated.

- **Tasks (Sub-Activities):**

Each phase of the model consists of several tasks or sub-activities.

- **Requirement Engineering Sub-Activities:**

Requirement engineering phase consists of the following sub activities:

- Requirement Elicitation
- Requirement Analysis
- Requirement Specification
- Requirement Testing
- Requirement Change Management

- **Design Sub-Activities:** Design phase has following sub activities:

- Architectural Design
- Conceptual Design
- Presentational Design
- Navigational Design

- **Implementation Sub-Activities:** Implementation phase has following activities:

- Coding
- Unit Testing
- Integration testing
- Heuristic Evaluation
- Deployment

The sketch of the whole model is given in Figure 2.

- **Artifacts**

An artifact is something that is produced, modified, or used by a task. Roles are responsible for creating and updating artifacts. Artifacts are subject to version control throughout the project lifecycle. Following are some of the artifacts that should be produced during the development of the project using the proposed approach:

- **Requirement Draft:** Containing the initial requirements given by the users.
- **Software Requirement Specification:** Containing the detail description of the requirements after analysis.
- **Requirement Testing Document:** Containing the test cases for checking the requirements.
- **Design Document:** Containing the architectural, conceptual, navigational and presentational design of the software to be produced.
- **Code Document:** Containing the code of the software.
- **Heuristic Evaluation Document:** Containing the reviews of the evaluators about the interface of the web application.
- **Test Document:** Containing the unit and integration test plans for the web applications.

### c. Process Content

#### • Process (Working of the Model)

Proposed usability driven web based software model starts with the requirement engineering phase. Requirement Engineering is the iterative phase containing requirement elicitation, requirement analysis, requirement specification, requirement validation, requirement testing, and requirement change management. The phase starts with the requirement elicitation activity and ends up with the requirement change management. In the whole process new requirements can arrive at any stage which needs to be accommodated. Phases of the proposed model are there in the coming section.

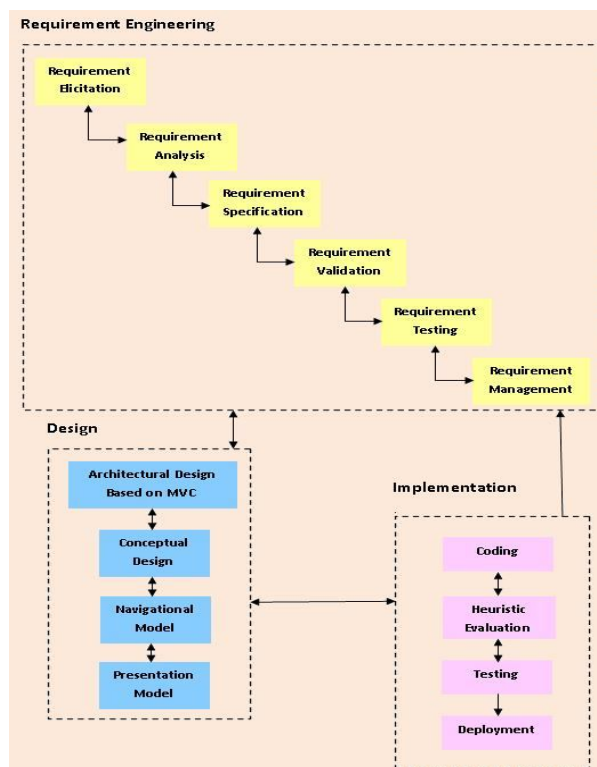


Figure 1: Proposed Usability Driven Web based Software Process Model

#### • Model Phases:

##### Phase-1: Requirement Engineering

Main requirement engineering process consists of several phases. Detailed description of each phase is given here as under:

**Requirement elicitation** involves extracting the right information from the stakeholders and users. Information can be gathered from several sources such as by documents, legacy applications, interviews etc. In this phase all kind of requirements such as Functional and Non functional requirements are gathered. As usability is the main concern in web application development, in particular the user interface requirements are also be focused and gathered.

The goal of the **requirement analysis** activity is to evaluate the requirements gathered in Requirement Elicitation activity. The requirements are analyzed on the basis of several properties i.e. clarity, completeness, contradiction, ambiguity etc [12]. The activity involves understanding the relationship among different requirements and shaping those relationships to achieve a successful result. In this activity, requirements are divided into three categories:

- a) Functional Requirements
- b) Non-Functional Requirements
- c) Usability Requirements

Then, analysts read the requirement draft, highlight the problems in Functional, Non functional and usability requirements. They have to pass the requirement draft through several stages such as Necessity checking, Consistency and Completeness check, and feasibility checking. In necessity checking analyst has to determine the need for the requirements, whether they are contributing to the goal of the business or not. Consistency and Completeness checking involves to make it sure that no requirements are contradictory and are not incomplete or missed out. While in feasibility checking it is ensured that the requirements are feasible in the context of budget, schedule and technology. Here, model considered Usability requirements as separate entity due to the fact the Usability is the core and important consideration in the web development environment. After analysis the analysts will come up with errors and problems in the requirements such as infeasible, inconsistent, incomplete and unnecessary requirements, which are then negotiated with the customer to come up with the requirements which are free of these problems.

Next activity in the requirement engineering phase is to come up with a complete **Software requirement specification**. In this activity analyst has to document all requirements (Functional, Non functional and Usability). SRS should be documented according to the world wide accepted standards so that it should clearly state all the requirements of the product.

The basic objective of **requirement validation** activity is to ensure that the SRS should imitate the actual and crucial requirements accurately and clearly. In this activity, SRS is inspected and reviewed by the customer to find errors and other matters of concern in the requirements of the system. In this review it is only validated whether the requirements are correct or not, other factors such as quality, readability, testability and user interface in particular are also considered. If requirements are reviewed and validated, not only a substantial fraction of the errors are detected by them, but a vast majority of the remaining errors are detected soon afterward in the design activity [13].

Next sub activity requirement testing involves the tests cases to be produced to check the requirements whether they are correct, complete, consistent, measurable, quantifiable, traceable and current.

Requirement management is the next sub activity in the proposed web development model. Change Management

process starts with the change request i.e. when a request for any requirement arrives. The request may consist of new requirements or can be any change to the existing requirements. The requirement then can be analyzed on the basis of its impact, whether it is feasible to the system in its each and every respect or not.

After impact analysis if the requirement under consideration is found to be feasible, is approved and placed in the requirement document with the rationale for the change. If the requirement is not feasible it is negotiated with the customer, then is approved or discarded.

## Phase-2: Design

**Design** phase of the proposed web development model is given below:

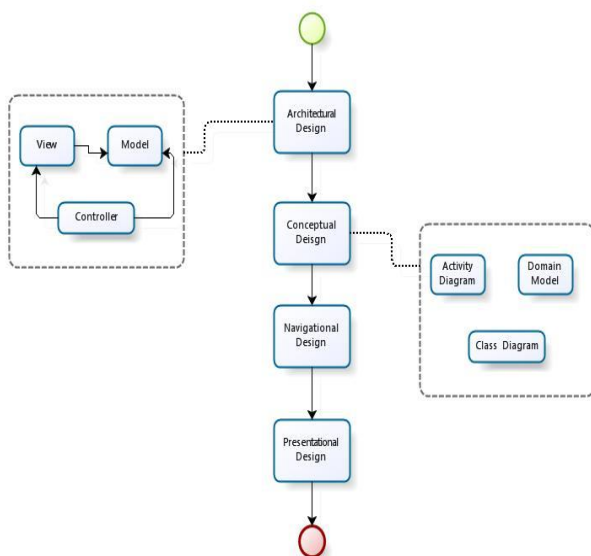


Figure 2: Design Phase of Proposed Model

Phase starts by designing the **architectural model** of the web application to be developed. The architectural design is based on Model View Controller [14] design pattern.

**Conceptual design** consists of various models such as

- **Domain Model:** Domain Model represents the basic entities and their relationships, and also encapsulates the business logic.
- **Sequence Diagram:** are constructed to show the overall workflow of the system. They can be used to describe the business and operational step-by-step workflows of components in a system.
- **Class Diagram:** In the conceptual design of a system, a number of classes are identified and grouped together in a class diagram which helps to determine the static relations between them. Class Diagram shows different classes, their attributes and functions also.

While modeling web application, the non-linearity of hypertext is the most important consideration taken into account. Thus, it is very important to design hypertext structure. So, the next step in web process model is to design the **navigational** or **hypertext model**. The objective of the hypertext model also called as navigational model is

to specify the navigability through the content of web application i.e., the navigation path available to the user.

Next model that we design in our web application development is the **presentation model**. It aims at designing the structure and behaviour of the user interface. In addition, the communication and representation task of the Web application are taken into account.

## Phase-3: Implementation

After designing, it is time to **implement** what we have designed so far. The phase (Figure 3) evolves with the actual coding of the web application and afterwards conducting the tests. The implementation is divided into following activities:

**Coding** involves implementation of classes and objects involved in the system. The coding is done in units and then these units are integrated to form a whole working system.

**Heuristic Evaluation** is the method of usability analysis where a number of evaluators are presented with user interface design and are asked to comment on it. They tried to come up with the opinion what is good and bad in the interface. Heuristic evaluation is usually carried out by the small set of evaluators. They evaluate the user interface design and judge its compliance with recognized usability principles which are commonly called as heuristics [15]. The result of this evaluation will be the related usability issues or problems with references to those usability principles that were violated by the design in each case in the opinion of the evaluator.

Next activity is the **unit testing**. As the system is developed in units, so unit testing will play an important role to find the bugs in the system. The primary goal of the unit testing is to take the smallest piece of software and determine whether it is behaving exactly as we expect or not.

**Integration testing** involves combining the individual modules and then testing them as a group. It involves taking the different modules as input which are unit tested, groups them in larger aggregates, applies tests defined in an integration test plan to those aggregates, and delivers as its output the integrated system.

The purpose of the **deployment** is to successfully produce product releases, and deliver the software to its end users.

Following tables (Table 1, Table 2 and Table 3) portrays the relationship between the method content and process content of proposed model:



Table 1: Process Content with their corresponding Method Content (Roles)

	Method Content (Roles)					
	Stakeholder	System Analyst	Designer	Developer	Tester	Project Manager
Process Content	Req. Elicitation	x	x			
	Req. Analysis		x			
	Req. Specs		x			
	Req. Validation	x				
	Req. Testing				x	
	Req. Management		x			
Process Content	Architectural Design			x		
	Conceptual Design			x		
	Navigational Design			x		
	Presentational Design			x		
Process Content	Coding			x		
	Heuristic Evaluation				x	
	Unit Testing				x	
	Integration Testing				x	
	Deployment			x		

Table 2: Process Content with their corresponding Method Content (Artifacts)

		Method Content (Artifacts)						
		Req. Draft	SRS	Req. Test Doc	Design Doc	Code	Heuristic Eval. Doc	Test Plan
Process Content	Req. Elicitation	x						
	Req. Analysis	x						
	Req. Specs		x					
	Req. Validation		x					
	Req. Testing			x				
	Req. Management		x					
	Architectural Design				x			
	Conceptual Design				x			
	Navigational Design				x			
	Presentational Design				x			
	Coding					x		
	Heuristic Evaluation						x	
	Unit Testing							x
	Integration Testing							x
	Deployment							

Table 3: Process Content with their corresponding Method Content (Disciplines)

	Method Content (Disciplines)		
	Req. Engineering	Design	Implementation
Process Content	Req. Elicitation	x	
	Req. Analysis	x	
	Req. Specs	x	
	Req. Validation	x	
	Req. Testing	x	
	Req. Management	x	
Process Content	Architectural Design		x
	Conceptual Design		x
	Navigational Design		x
	Presentational Design		x
Process Content	Coding		x
	Heuristic Evaluation		x
	Unit Testing		x
	Integration Testing		x
	Deployment		x

### III. CONCLUSION

Proposed usability driven web based software process model is meant for delivering the product in three simple phases. The main feature of the model includes its flexibility i.e. the movement from any phase to another. The model does not freeze any phase, one can easily move from design to requirement engineering phase and from implementation to requirement engineering and design. A flexible and structured approach of web development is followed that emphasizes the continuous delivery of quality software that is valuable to stakeholders.

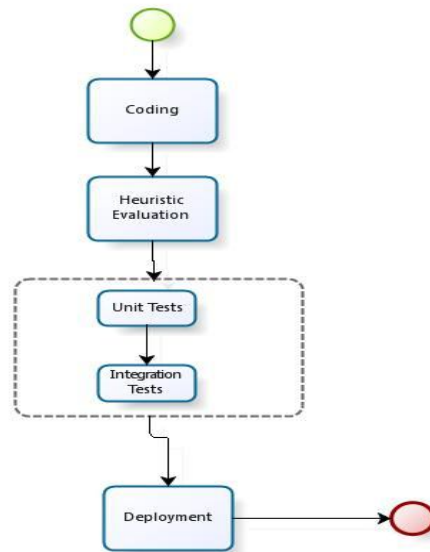


Figure 1: Implementation Phase

### REFERENCES

- [1] Sommerville I. Software Engineering, China: Addison-Wesley 2006 8th Edition. pp 08.
- [2] Madachy R.J., Software Process Dynamics, New Jersey: IEEE Press, John Wiley & Sons, 2007, pp 07.
- [3] Humphrey W., "Comments on eXtreme Programming", eXtreme Programming Pros and Cons: What Questions Remain", IEEE Computer Society Dynabook. September 2001, Available at: <http://www.computer.org/SEweb/Dynabook/HumphreyCom.htm>
- [4] Capretz L. F. "Y: A New Component-Based Software Life Cycle Model". Journal of Computer Science 2005.
- [5] V-Model (Software Development), Internet: [http://en.wikipedia.org/wiki/V-Model\\_\(software\\_development\)](http://en.wikipedia.org/wiki/V-Model_(software_development)) [May 21, 2010].

- [6] Sommerville I. *Software Engineering*, China: Addison-Wesley 2006 8th Edition. pp 71
- [7] Jalote P., "Chapter 2: Software Processes" in A Concise Introduction to Software Engineering, London: Springer 2008, pp 17-30.
- [8] Pressman R. S., "Chapter 02: The Process" Software Engineering, A Practitioner Approach", 5th ed. New York: McGraw-Hill, 2001, pp 31-35.
- [9] B.W. Boehm, "A Spiral Model for Software Development and Enhancement", ACM SIGSOFT Software Engineering Notes. August 1986, vol. 11 pp 14-24.
- [10] Juric R. "Extreme Programming and its development practices". IEEE 22<sup>nd</sup> International Conference on Information Technology. June 2000 pp 97-104.
- [11] Sommerville I. *Software Engineering*, China: Addison-Wesley 2006 8th Edition. pp 398
- [12] Requirement Analysis (2010),  
[http://en.wikipedia.org/wiki/Requirements\\_analysis](http://en.wikipedia.org/wiki/Requirements_analysis),  
[June 22, 2010]
- [13] Jalote P., "Chapter 3: Software Requirement Analysis and Specification" in A Concise Introduction to Software Engineering, London: Springer 2008, pp 64.
- [14] Model-View-Controller (2010),  
<http://en.wikipedia.org/wiki/Model-View-Controller>.  
[June 22, 2010].
- [15] Nielsen, J. and Molich, R. "Heuristic Evaluation of User Interfaces ", ACM CHI'90 Conference on Human Factors in Computing Systems, 1990 pp. 249-256.

# Power-Based Key Hopping (PBKH) and Associated Hardware Implementation

Rabie A. Mahmoud<sup>1</sup>, Magdy Saeb<sup>2</sup>

1. Department of Mathematics, Faculty of Science, Cairo University, Cairo, Egypt.
2. Computer Engineering Department, Arab Academy for Science, Tech. & Maritime Transport, Alexandria, Egypt,  
[mail@magdysaeb.net](mailto:mail@magdysaeb.net)

**Abstract:** Power-Based Key Hopping (PBKH) is a process of key hopping that is used to interchange the user key of a cipher. Power based key hopping is founded on the idea of dynamic frequency power-based hopping to change the user key. This is achieved through computing the power of a previous cipher packet and comparing it with a standard value. In this work, we discuss various key hopping methods and suggest a procedure of power based key hopping. Moreover, we provide a Field Programmable Gate Array (FPGA) hardware implementation of the proposed key hopping technique.

**Keywords:** Power Based Key Hopping; Security; Hardware; FPGA

## I. INTRODUCTION

Power-Based Key Hopping (PBKH) is a key hopping method that is based on utilizing the idea of power based dynamic frequency hopping to provide a new procedure of key hopping. PBKH utilizes four keys; one acts as the authentication key where the power of cipher text packet will be the standard of changing the key. In the following sections, we provide a review of various key hopping methods, the structure of power based key hopping and the formal description of power based key hopping algorithm. Moreover, we provide the details of our hardware implementation, a discussion of the results of the FPGA implementation and finally a summary and our conclusions.

## II. GENERAL KEY HOPPING TECHNIQUES

Various key hopping techniques were suggested in the literature. The major procedures are summarized as follows:

### A. NextComm Key Hopping

“NextComm” [1] has proposed changing the keys frequently to overcome the weakness of Wired Equivalent Privacy (WEP) without incurring possible overheads and other complexities of the network. The key hopping process is designed as follows:

- The shared keys which are maintained in network devices are not used directly for encryption/decryption but the session keys are derived using a robust hash function for encryption

and decryption purposes where the 128-bit MD5 one-way hash function is chosen.

- The session seeds which are the output of the large counter are changed frequently to ensure that the resulting session keys not to be reused where the size of the counter should also be large enough to ensure the seeds are not reprocessed during the lifetime of the secret shared keys.
- Session keys are used in the same way in encryption and decryption as the shared keys are used in the standard WEP process.

### B. State Based Key Hop (SBKH)

State Based Key Hop (SBKH) protocol is created to provide a strong lightweight encryption scheme for battery operated 802.11 devices. SBKH includes base key pair, key duration, RC4 states, offset and an explicit SBKH sequence counter as the sharing parameters between communicating entities. Base key pair consists of two 128-bit keys which are shared by all entities in a network. The keys can be used as per-session keys that are agreed upon between the two communicating nodes. SBKH uses a 24-bit initialization vector IV within the original 802.11 WEP data frames as a SBKH sequence counter [2].

### C. Dynamic Re-Keying with Key Hopping (DRKH)

Dynamic Re-keying with Key Hopping (DRKH) encryption protocol uses RC4 encryption technique where each two communicating entities share a secret list of the static keys. One of these keys is an authentication key that is used to encrypt and decrypt mutual authentication messages. The transmitted packets in DRKH between the two communicating nodes take place in several consecutive sessions where the access point (AP) in WLANs will be responsible for determining the sessions' start and end times. The access point maintains the session counter. A one-way hash function such as MD5 or SHA-1 is used in DRKH to hash the session counter with each of the four secret keys to generate four session keys and then these session keys are used instead of the secret keys to encrypt and decrypt packets during the lifetime of a session. Likewise, an initialization Vector (IV) is used in DRKH where each session key has a corresponding IV which value is incremented for every new packet to be

encrypted using that particular session key. Non-linear lookup-table based substitution technique to mix a session key with its corresponding IV value is used to reinitialize RC4 state instead of using the key scheduling algorithm where one of the four derived session keys is selected according to a previously agreed upon key hopping sequence for every packet to be encrypted. The Key Hopping Sequence determines the order in which session keys are used to encrypt and decrypt packets where the AP sends the key hopping sequence to the wireless station after the completion of a successful mutual authentication process [3].

#### D. Dynamic frequency hopping based power (DFHBP)

These methods can be generally divided into three categories:

- Full-replacement dynamic frequency hopping which assumes changing all current frequencies of poor quality after each physical layer frame,
- Worst dwell (DFH) where only one frequency (the lowest quality one) in a frequency-hop pattern is changed,
- Threshold (SIR)-based DFH where a subset of currently used frequencies is changed,

More recently, another version of dynamic frequency hopping is called Dynamic Frequency Hopping Based Power [4] which can be explained as follows:

A subset of currently used frequencies is changed depending on a criterion value of power. In each frame, the power is measured on the used frequencies and the current hopping pattern is changed if the measured power does not achieve the required threshold. Only the frequencies in with low powers are changed. Any frequency that meets the threshold can be used as a replacement frequency, and thus there is no need to scan all possible frequencies.

### III. POWER BASED KEY HOPPING (PBKH)

Our proposed method uses a discrete form of signal analysis where the signal power is measured as the mean of the signal encountered. In this case, the power of discrete signal  $x_{[i]}$  with length  $L$  is determined by the mean of  $x_{[i]}$

$$Pr = \frac{1}{L} \sum_{i=1}^L |x_{[i]}|^2 = \text{mean}(x^2) \quad (1)$$

In other words, the power of a packet of binary sequence with length  $L$  can be determined by the number of ones in that packet compared with the number of zeros [5]. In this respect, our proposed power based key hopping (PBKH) has its roots in the above mentioned Dynamic frequency hopping based power (DFHBP). The proposed method utilizes four keys. Each two communicating entities maintain four secret keys, one of them will be an authentication key and the first key will be used as the default key at each time the communication session is started. The first key will be used to encrypt and decrypt the first plaintext packet then the sender and receiver

compute the power of the ciphertext to determine which key will be used to encrypt and decrypt the next packet. When the power of ciphertext packet which is the number of ones equals a certain predefined value known before hand between the communicating entities, the same key will be used to encrypt and decrypt the next packet. If the power of ciphertext packet is larger than this value, then the next key from the set of secret keys is to be used. However, the previous key from secret key list is to be used to encrypt and decrypt the next packet when the power of ciphertext is less than this certain value. Initialization vector IV will be concatenated to the keys and incremented by one each clock independent of the used key where long IV consisting of 64 bits. In this process, the PBKH does not specify a key duration between the two communicating entities. Therefore, the process provides a higher degree of randomness increasing the entropy and thus improving the security.

### IV. ALGORITHM

In this section, we provide the formal description of the power based key hopping algorithm as follows:

#### **Algorithm: Power Based Key Hopping (PBKH)**

**INPUT:** Plain text message ( $P$ ), User Authentication Key ( $K$ ), User Key ( $K1$ ), User Key ( $K2$ ), User Key ( $K3$ ), Initialization Vector ( $IV$ ), predefined and known value between communicating entities ( $a$ ).

**OUTPUT:** Cipher Text ( $C$ )

**Algorithm body:**

**Begin**

**Begin power based key hopping**

1. Read user keys;
2. Authentication key with IV is used for authentication messages.
3. Encrypt first packet by calling the encrypt function using the user key  $K1$  (default key) concatenating initial vector  $IV$ ;
4. Compute the power of the encrypted packet;
5. Compare the value of power of encrypted packet with certain value  $a$  defined and known between the communicating nodes;

5.1 If the power  $Pr = a$  then the same key is used to encrypt/decrypt the next packet.

5.2 If the power  $Pr > a$  then the next key is used to encrypt/decrypt the next packet.

5.3 If the power  $Pr < a$  then the previous key is used to encrypt/decrypt the next packet

6. Increment the initial vector  $IV$  by one  $IV = IV + 1$ ;

7. Encrypt the packet calling the encrypt function using the chosen key (in step 5) concatenating with new initial vector  $IV$

8. Repeat steps 4 till 7 if message cache is not empty;

9. When the plain text messages are finished then halt;

**End power based key hopping;**

**End Algorithm;**

**Function ENCRYPT**

**Begin**

1. Read next message bit;
  2. Read next key bit from user key;
  3. Use any standard encryption algorithm to encrypt the plaintext; (In the very simple version of this protocol, one can use XOR operation, however generally this is not recommended)
  4. Perform the encryption;
  5. Store the resulting cipher;
- End;**

This procedure can be summarized as shown in Figure 1. In this Figure, we use the state diagram for a 64-bit IV and 192-bit user keys to encrypt and decrypt 256-bit plaintext packets with a given power threshold value equal to 128.

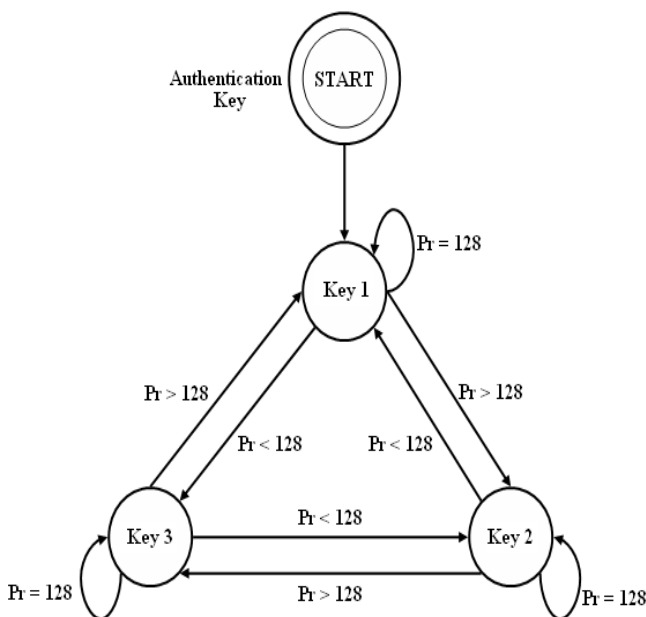


Figure 1. Power based key hopping scheme

## V. FPGA IMPLEMENTATION

The concept of the power based key hopping applied to change cipher key leads to a relatively easy to design FPGA-based implementation. We have implemented the proposed technique applying VHDL hardware description language [6], [7], [8] and utilizing Altera design environment Quartus II 9.1 Service Pack 2 Web Edition [9]. The circuit is based on the idea of encrypting 256-bit plaintext blocks using 256-bit user keys that produce 256-bit ciphertext blocks. Four 192-bit keys are used and a 64-bit initialization vector IV which is incremented by one at each falling-edge of the clock trigger. Authentication key and three user keys will be interchanged depending on the power of previous encrypted text packet where the power threshold that is to be compared with is equal to 128. The schematic diagram for a demonstrative 256-bit power based key hopping module is shown in Figure 2. The design was

implemented using an EP2C70F896C6, Cyclone II family device.

The worst case pin-to-pin delay was found to be equal to 15.592 ns from source pin "INPUT [110]" to destination pin "OUTPUT [110]". The longest pin-to-register delay was equal to 23.745 ns and the shortest pin-to-register delay was 9.485 ns. The longest register-to-pin delay was 9.206 ns and the longest register-to-register delay was 17.064 ns. The worst case clock-to-output delay was 12.359 ns from clock "CLK2" to destination pin "OUTPUT [69]" through register "KH\_Power: inst|pre\_temp\_key [26]". Clock "CLK2" had an internal  $f_{max}$  of 57.91 MHz (period = 17.267 ns) between source register "KH\_Power:inst|IV [60]" and the destination register "KH\_Power:inst|pre\_temp\_key [160]". A series of screen-captures of the different design environment output are shown in Figures 3 to 8. Figures 3, 4, 5, and 6 provide the indication of a successful compilation and parts of RTL for power based key hopping respectively. These are shown as a series of repeated MUX's with different connections and a state machine block. Figure 7 displays the power based key hopping simulation showing the output encrypted bits. Figure 8 demonstrates the floor plan. The details of the analysis and synthesis report are shown in appendix A.

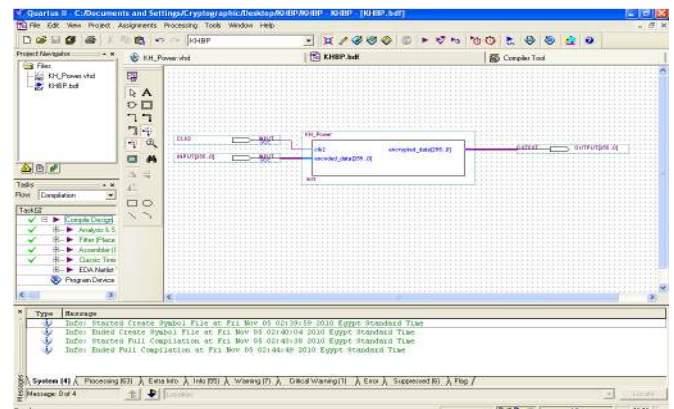


Figure 2. Schematic diagram of power based key hopping implementation

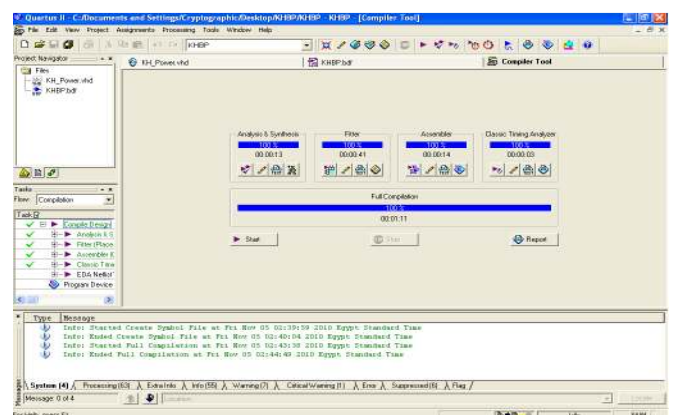


Figure 3. Compiler tool screen showing correct implementation



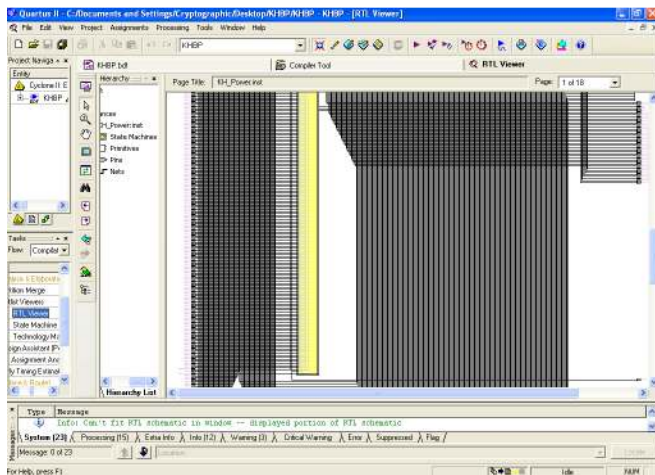


Figure 4. RTL screen for part of power based key hopping implementation

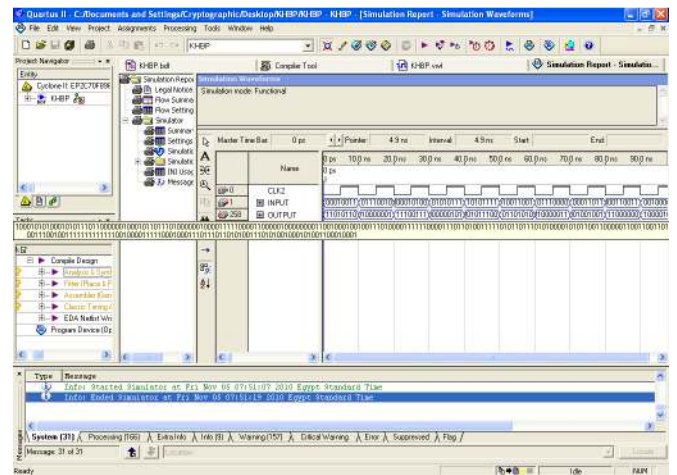


Figure 7. Simulator screen showing the output encrypted data depending on power based key hopping

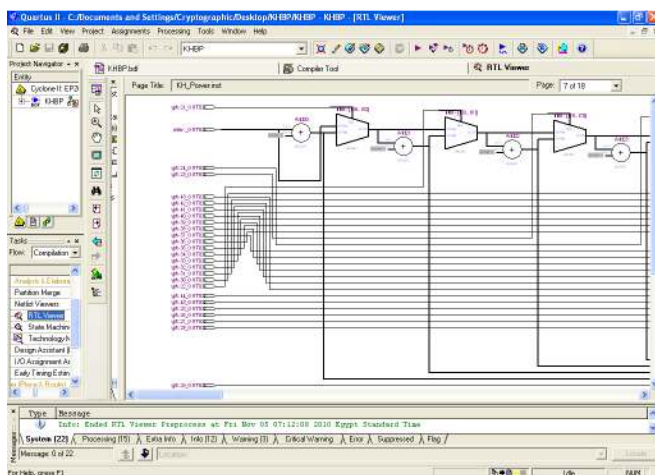


Figure 5. RTL screen for part of the resulting circuit diagram

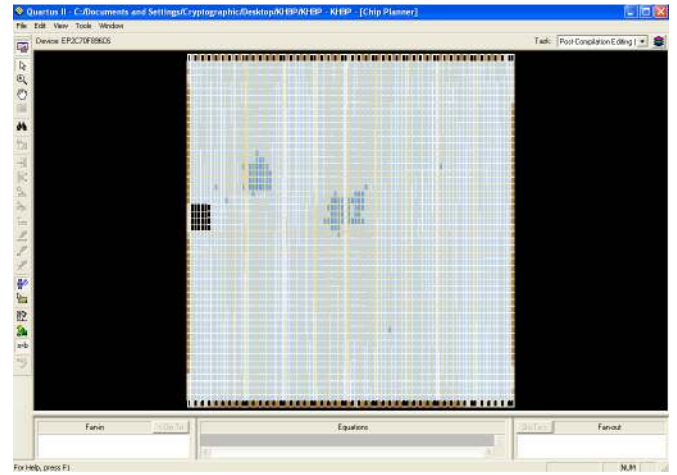


Figure 8. Floor-plan of power based key hopping implementation

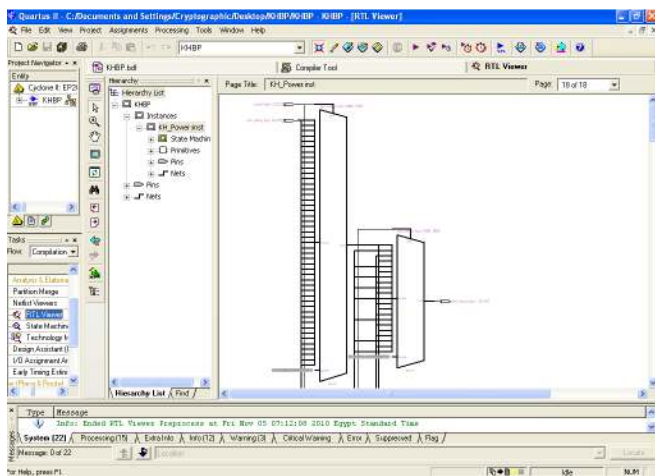


Figure 6. RTL screen for part of power based key hopping implementation

## VI. CONCLUSION

We have provided a brief discussion of the concept of power based key hopping and its hardware implementation. The details of the proposed procedure were discussed in sections iii and iv. The method uses a discrete form of signal analysis where the signal power is measured as the mean of the signal encountered. The power of a packet of binary sequence with a finite length can be determined by the number of ones in that packet. The proposed method utilizes four keys. Each two communicating entities maintain four secret keys, one of them will be an authentication key and the first key will be used as the default key at each time the communication session is started. When the power of ciphertext packet computed based on the number of ones equals a certain predefined value between the communicating entities, the same key will be used to encrypt and decrypt the next packet. Otherwise the key is changed to the next key of the key set.

The hardware implementation uses a set of modules that were carried out applying VHDL and then joined together using the schematic editor. The resulting circuit provides a proof-of-concept FPGA implementation. It was shown that the worst case pin-to-pin delay is equal to 15.592 ns. Moreover, area and speed optimization were performed and it is shown that the worst case pin-to-pin delay is equal to 15.645 ns in the case of area optimization and 15.383 ns in speed optimization. Moreover, high fan-out reduces the usage of global interconnection resources. Therefore, the speed optimization decreases the use of global resources. This is clearly demonstrated in the synthesis report of our design. A comparison with other implementations is not relevant since this is the first time this power based key hopping is FPGA-implemented. This and other related effects will be dealt with in future development of the device.

## REFERENCES

- [1] W. Ying, "Key Hopping – A Security Enhancement Scheme for IEEE 802.11 WEP Standards," NextComm Inc, USA, 2002.
- [2] K. Srinivasan, and S. Mitchell, "State Based Key Hop (SBKH) Protocol," Sixteenth International Conference on Wireless Communications Wireless 2004, Alberta, Canada, 2004.
- [3] A. M. Kholaf, M. B. Fayek, H. S. Eissa, and H. A. Baraka, "DRKH: Dynamic Re-Keying with Key Hopping," McMaster University, Hamilton, Canada, 2005.
- [4] H. Bli, and H. Salim, "Performance Enhancement of GSM Cellular Phone Network using Dynamic Frequency Hopping," Engineering & Technology Journal, Vol. 26, No.3 (2008), University of Technology, Baghdad, Iraq, 2008.
- [5] V. Saxena, "K-Delta-1-Sigma Modulators for Wideband Analog-to-Digital Conversion," Electrical and Computer Engineering Department, Boise State University, Idaho, USA, 2010.
- [6] P. P. Chu, RTL Hardware Design Using VHDL, John Wiley & Sons, Inc., New Jersey, 2006.
- [7] E. O. Hwang, Digital Logic and Microprocessor Design with VHDL, La Sierra University, Riverside, California, USA, 2005.
- [8] V. A. Pedroni, Circuit Design with VHDL, MIT Press, 2004.
- [9] Altera's user-support site:  
<http://www.altera.com/support/examples/vhdl/vhdl.html>

## APPENDIX A

### The analysis and synthesis report details

Family: Cyclone II  
Device: EP2C70F896C6  
Total logic elements: 887 out of 68,416 (1%)  
Total combinational functions: 887  
Logic element usage by number of LUT inputs  
    -- 4 input functions: 221  
    -- 3 input functions: 249  
    -- <=2 input functions: 417  
Total registers: 80 out of 70,234 (< 1%)  
Total pins: 513 out of 622 (82 %)  
Total memory bits: 0 out of 1,152,000 (0 %)  
Embedded Multiplier 9-bit elements: 0 out of 300 (0 %)  
Total PLLs: 0 out of 4 (0 %)

Optimization Technique: Balanced  
Maximum fan-out: 80  
Total fan-out: 2721  
Average fan-out: 1.83

### Fitter Summary

Block interconnects: 1206 out of 197,592 (< 1%)  
C16 interconnects: 387 out of 6,270 (6 %)  
C4 interconnects: 1220 out of 123,120 (< 1%)  
Direct links: 233 out of 197,592 (< 1%)  
Global clocks: 1 out of 16 (6 %)  
Local interconnects: 435 out of 68,416 (< 1%)  
R24 interconnects: 511 out of 5,926 (9 %)  
R4 interconnects: 1394 out of 167,484 (< 1%)  
Nominal Core Voltage: 1.20 V  
Low Junction Temperature: 0 °C  
High Junction Temperature: 85 °C

The usage number of logic elements and their connections in the device can be changed depending on the optimization technique which is used for synthesizing the power based key hopping. Table A1 and table A2 show the number of usage logic elements and the interconnections between them in Area, Speed, and Balanced optimization technique. We noticed that the total number of usage logic elements in the device increased in speed optimization technique comparing with balanced and area optimization techniques improving the routability of the design through using more C4 interconnects, which are considered the main interconnects in the device, and applied 1.86 average fan-out. While area and balanced optimization techniques give less average fan-out 1.83.

Table A1. A synthesis comparison between optimization technique implementations of power based key hopping

	Balance	Area	Speed
Total logic elements	887	895	909
Total combinational functions	887	895	909
4 input functions	221	215	228
3 input functions	249	259	262
<=2 input functions	417	421	419
Total fan-out	2721	2735	2796
Average fan-out	1.83	1.83	1.86

Table A2. A fitter comparison between optimization technique implementations of power based key hopping

	Balance	Area	Speed
Block interconnects	1206	1201	1211
C16 interconnects	387	390	354
C4 interconnects	1220	1226	1291
Direct links	233	187	213
Global clocks	1	1	1
Local interconnects	435	451	451
R24 interconnects	511	615	550
R4 interconnects	1394	1439	1373

The delays comparison between optimization techniques was extracted from the timing reports of implementing area and speed optimization. Figure A.1 shows a comparison chart between various implementation delays.

● In area optimization

- The worst case pin-to-pin delay was found to be equal to 15.645 ns from source pin "INPUT[23]" to destination pin "OUTPUT[23]".
- The longest pin-to-register delay was 23.955 ns and the shortest pin-to-register delay was 8.878 ns. The longest register-to-pin delay was 9.705 ns and the longest register-to-register delay was 16.611 ns.
- The worst case clock-to-output delay was 12.810 ns from clock "CLK2" to destination pin "OUTPUT[116]" through register "KH\_Power:inst|pre\_temp\_key[26]".
- Clock "CLK2" had internal fmax of 59.51 MHz (period = 16.803 ns) between source register "KH\_Power:inst|IV[52]" and destination register "KH\_Power:inst|pre\_temp\_key[161]".

● In speed optimization

- The worst case pin-to-pin delay was found to be equal to 15.383 ns from source pin "INPUT[159]" to destination pin "OUTPUT[159]".
- The longest pin-to-register delay was 24.002 ns and the shortest pin-to-register delay was 9.566 ns. The longest register-to-pin delay was 8.687 ns and the longest register-to-register delay was 17.476 ns.
- The worst case clock-to-output delay was 11.770 ns from clock "CLK2" to destination pin "OUTPUT[104]" through register "KH\_Power:inst|pre\_temp\_key[26]".
- Clock "CLK2" had internal fmax of 56.37 MHz (period = 17.74 ns) between source register "KH\_Power:inst|IV[38]" and destination register "KH\_Power:inst|pre\_temp\_key[30]".

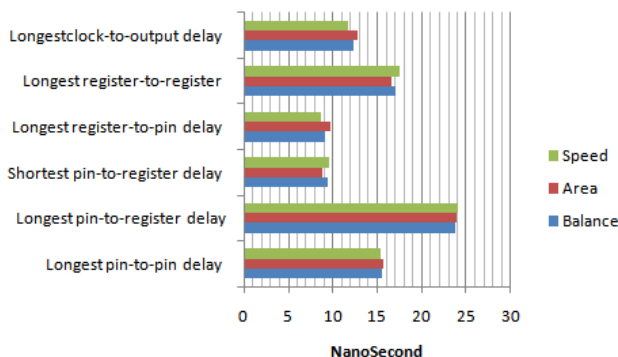


Figure A.1. Delays in our design of power based key hopping implementation

APPENDIX B

Sample VHDL code for 8-bit Power Based Key Hopping module

```
LIBRARY ieee;
USE ieee.std_logic_1164.all;
USE ieee.std_logic_arith.all;
USE ieee.std_logic_unsigned.all;

ENTITY KH_Power IS
    port(clk2 : in std_logic;
          encoded_data : in std_logic_vector (7 downto 0);
          encrypted_data : out std_logic_vector (7 downto 0);
          key : out std_logic_vector (7 downto 0));
END KH_Power ;

ARCHITECTURE behavioral OF KH_Power IS
    signal IV : std_logic_vector (1 downto 0) := "00";
    signal key_authentication, pre_temp_key :
        std_logic_vector (5 downto 0) := "110011";
    signal key1 : std_logic_vector (5 downto 0) := "010101";
    signal key2 : std_logic_vector (5 downto 0) := "111000";
    signal key3 : std_logic_vector (5 downto 0) := "101010";
    signal temp_key : std_logic_vector (7 downto 0);
    signal ypt : std_logic_vector (7 downto 0);

    type state_type is (S1, S2, S3);
    signal state : state_type := S1;

BEGIN
    temp_key <= IV & pre_temp_key;
    process (clk2, encoded_data, state, temp_key)
        variable num : integer range 0 to 8;
    begin
        for i in 7 downto 0 loop
            ypt(i) <= encoded_data(i) xor temp_key(i);
        end loop;

        if falling_edge(clk2) then
            num := 0;
            for j in 7 downto 0 loop
                if ypt(j) = '1' then num := num + 1;
            end if;
            end loop;

            case state is
                when S1 => if num = 4 then
                    state <= S1; pre_temp_key <= key1;
                elsif num < 4 then
                    state <= S2; pre_temp_key <= key2;
                elsif num > 4 then
                    state <= S3; pre_temp_key <= key3;
                end if;
                when S2 => if num = 4 then
                    state <= S2; pre_temp_key <= key2;
                elsif num < 4 then
                    state <= S3; pre_temp_key <= key3;
                elsif num > 4 then
                    state <= S1; pre_temp_key <= key1;
```

```

end if;

when S3 => if num = 4 then
    state <= S3; pre_temp_key <= key3;
elsif num < 4 then
    state <= S1; pre_temp_key <= key1;
elsif num > 4 then
    state <= S2; pre_temp_key <= key2;
end if;

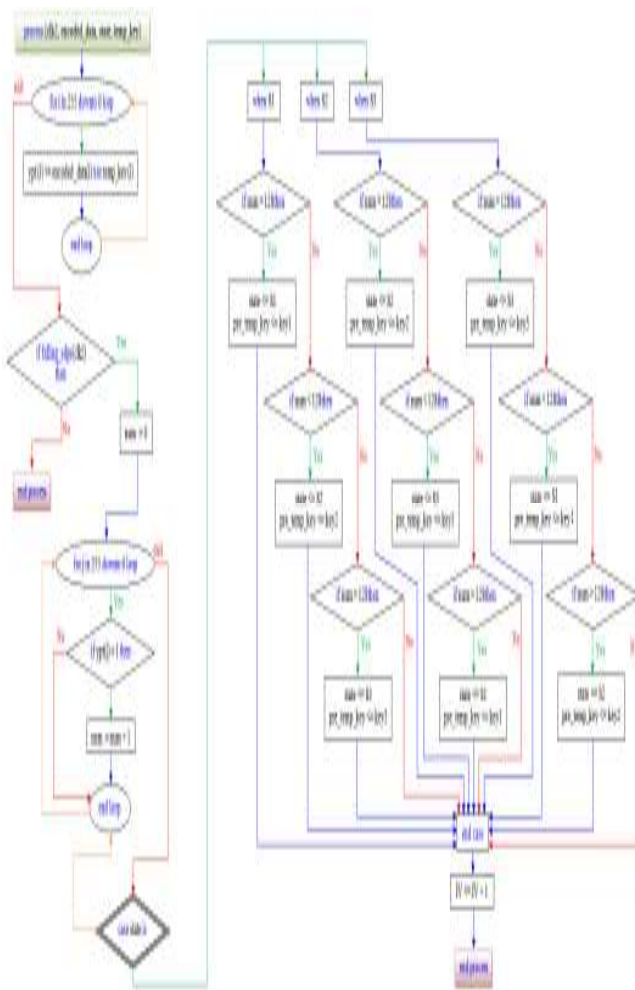
end case;
IV <= IV + 1;
end if;
end process;

temp_key <= IV & pre_temp_key;
encrypted_data <= ypt;
key <= temp_key;
END behavioral;

```

## APPENDIX C

VHDL Flowchart for 256-bit Power Based Key Hopping module



## AUTHORS' PROFILE



**Magdy Saeb** received the BSEE, School of Engineering, Cairo University, in 1974, the MSEE, and Ph.D. degrees in Electrical & Computer Engineering, University of California, Irvine, in 1981 and 1985, respectively. He was with Kaiser Aerospace and Electronics, Irvine California, and The Atomic Energy Establishment, Anshas, Egypt. Currently, he is a professor in the Department of Computer Engineering, Arab Academy for Science, Technology & Maritime Transport, Alexandria, Egypt; He was on-leave working as a principal researcher in the Malaysian Institute of Microelectronic Systems (MIMOS). His current research interests include Cryptography, FPGA Implementations of Cryptography and Steganography Data Security Techniques, Encryption Processors, Mobile Agent Security. [www.magdysaeb.net](http://www.magdysaeb.net).



**Rabie Mahmoud** received the BS. Degree, Faculty of Science, Tishreen University, Syria, in 2001, the MS. in Computational Science, Faculty of Science, Cairo University, Egypt in 2007. Currently, he is working on his Ph.D. Dissertation in the Department of Mathematics, Faculty of Science, Cairo University. His current interests include Image processing, Cryptography, FPGA Implementations of Cryptography and Data Security Techniques. [rabiemah@yahoo.com](mailto:rabiemah@yahoo.com)



# The Innovative Application of Multiple Correlation plane

Julaluk Watthananon

Faculty of Information Technology,  
King Mongkut's University of  
Technology North Bangkok,  
Bangkok, Thailand  
watthananon@hotmail.com

Sageemas Na Wichian

College of Industrial Technology,  
King Mongkut's University of  
Technology North Bangkok,  
Bangkok, Thailand  
sgm@kmutnb.ac.th

Anirach Mingkhwan

Faculty of Industrial and Technology  
Management, King Mongkut's University  
of Technology North Bangkok,  
Bangkok, Thailand  
anirach@ieee.org

**Abstract**—Presentation data with column graph and line graph is a well-known technique used in data explanation to compare and show direction that users can easily understand. However, the techniques has limitations on the data describing complex with multiple relations, that is, if the data contains diverse relationships and many variables, the efficiency of the presentation will decrease. In this paper, the mathematical method for multi relations based on Radar graph is proposed. The position of information approaches on the correlation plane referred to the distribution of content and the deep specific content. However, the proposed method analyzes the multi variants data by plotting in the correlation plane, and compared with the base line system. The result shows that the performance is higher than other methods in term of accuracy, time and features.

**Keywords**—Correlation plane; correlation boundary; correlation plot; Star plot; Radar graph

## I. INTRODUCTION

In statistics, bar graph and line graph are common types of graphs employed to explain data analyses, to compare directions and to represent a set of qualitative data with correlation between two variables [1]. Nonetheless, comparative analyses of more than two qualitative variables and multiple correlations have been increasingly implemented in many fields of work, namely weather conditions, context consistency of documents, etc. It is important to have a proper form of data presentation that can effectively send messages across to readers. One of the commonly used forms of data presentation is a radar chart that can represent data with correlation of over two variables in an effective manner due to its continuity and its ability to clearly compare many aspects of data plot correlations [2]. However, there are a number of limitations in presenting a larger amount of data with multiple correlations. Representatives of those relations need to be sought so as to determine appropriate data positions.

Generally, there are three methods of selecting representatives of data values with correlation of multiple variables. The three methods are as follows:

1) *Selecting the highest value*: classifying quantitative data of each variable, and then selecting the most quantities variables, for instance, in order to classify books categories [3], librarians will normally do on the essence of the books. Disadvantage of this method is other contents relating to other topics are decreased in the importance and deleted.

2) *Selecting from the mean*: By this method a value data representative from the mean or neutral value calculating from an outcome of added data divided by data amount. This method is usually employed in research to selecting variables representatives. However, it is not suitable for selecting data with multiple correlations because accurate data cannot be identified clearly.

3) *Calculating combined results of directions*: this is a highly successful technique commonly used with data with multiple variables [4], [5], [6]. A mathematic process is employed to acquire relation between rectangular and polar coordinates on a radar chart and proper coordinates' positions resulted from calculations of directions and distances of those relations. The authors name these plots **data correlation plots**. They are on correlation plane of connected lines and will confine the area, create an  $n$  axis and divide the plane within polar coordinates. The plane in this research is referred to as **the correlation plane**. The intersection of  $n$  axis is called the origin. Intersection of  $n$  axes will divide the plane into  $n$  parts. Each part is called a **correlation boundary**, details of which are elaborated in Section 3.

Hence, the authors have developed a concept of applying the method of calculating combined results of directions to present results in the correlation form as above mentioned definition. Furthermore, efficiency of presentation of implementing methods, directions and depth levels of the correlation to data with multiple variables was analyzed.

The rest of this paper is organized as follows: In the section 2 we provide a review of related works about star graph, polar coordinates, distance between plots, Dewey decimal classification and Dewey decimal classification – Multiple relations. Section 3, 4 and 5 present the definition of correlation such as: correlation plot, correlation plane and



correlation boundary, the concept of application and the experiments with the discussion on the results respectively. Section 6 is the conclusion.

## II. A REVIEW OF RELATED WORKS

Many researchers have studied and designed methods of presentation from the information retrieval format that allows users to access and easily understand with the visualization, such as in research Texas State Auditor's [1] presented how to use graphs representations of the relationships between two or more variables and the issues of interest. Yeh [2] presented star chart showing the target numeric variable among categories. The results showed the GRADAR procedure providing a new looks to the clinical data and helped checking the data, tables and clinical reports. Wang et al. [4], [5] proposed a new graphical representation for multi-dimensional data in multivariate. The experimental results showed the effectiveness of accurate classification. Klippel et al. [8] proposed that the best visual representations for a data set presented are: how to assign variables to rays and to add color to rays of a star plot graph. The results shown that the star plot graphs were meaningful; the represented data and star plot enhanced color had positive effects on the processing speed. Peng et al. [9] presented a new method for shape representation by converting the CSS descriptor circular vector map and defining two histograms in polar coordinate system. The advantages of their proposed are simplicity, execution speed and efficiency of well in clustering the shape images. Sukmar et al. [12] presented the construction of a new polygonal interpolant that was based on the concept of natural neighbors. They used technique to adapt the above construction on polygonal elements to quad tree meshes to obtain  $C^0(\Omega)$  admissible approximations along edges with "hanging nodes." Mohseni et al. [13] presented a method for treating the coordinate singularity whereby singular coordinates were redefined. Thus, the results showed the new pole treatment giving spectral convergence and more accurate for all. Demsar et al. [14] presented a new method for visualization "FreeViz". The results showed that the FreeViz was very fast and can presented high quality with clear class separation.

From the researches above, the most effective technique to present data was a compute of the relationships and presented a new method for intelligent visualization [4], [5], [9], [12], [14], [15], [16], [17] of data sets. In this paper, we also applied the star graph and polar coordinates to improve the classification correlation and presented the position of data. Since a normal plane cannot explain correlations of that calculated position as a result of the starting point originated from variables with multiple correlations. Below are theories of related works with techniques coming from these diverse fields.

### A. Star graph

The star graph (can call radar graph or spider graph) is a technique used to represent graphical data analysis with all variables in multivariate data sets. It consists of a different

variable of lines radiating from the center. It means the "data length" of each variable. The characteristics of radar graph are polygons under the frame of circle that shows many data in the same graph, so the principles of creation consist of:

1) *Determination of the axis*: Determination of the axis and number of axis displays data where we define one axis for one data, the first axes is in vertical (x-axis) and then circulates to the east. In addition, users can define the color [5], [8], weight and name of title.

2) *Plot the value on the axis*: Plot the value on the axis that starting from the origin (point O) to the circumference by assigning the position (x, y) on each axis.

### B. Polar coordinates

The polar coordinate [4], [5], [6], [9] is a popular method used to calculate the appropriate location of multi variances, in order to represent the data referred to multiple relations. The research of Wang et al. [4] shows that, this method can be classified of data efficiently. In previous works [6], we had analyzed and computed the correlation of document contents by DDC-MR method [3]. It showed that position could refer to the relationship of multiple variables effectively, so this paper we used the sum of vector method to represent the multi variances as shown in Figure 1.

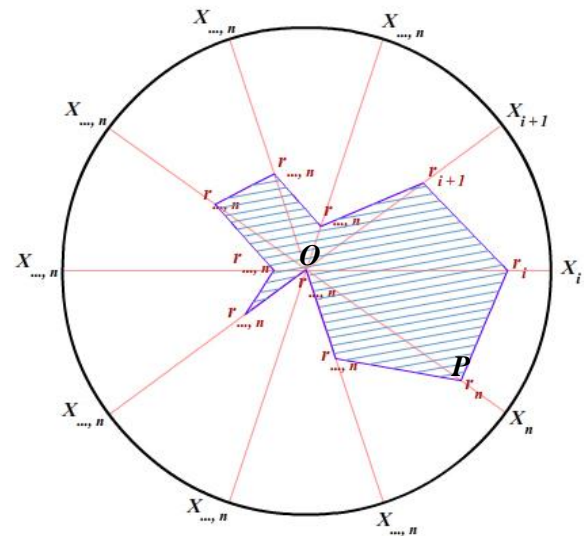


Figure 1. Example data with multi variances, where  $n$  is the number of variance,  $r_i$  is relationships between rectangular and polar Coordinates ( $r, \theta$ ).

In Figure 1 we show example data with multi variances, let  $r_{ij}$  denote the distance of a point  $P$  from the origin and the symbol  $O$  is the data length. The shade means area in the computed appropriate position of multiple variable, let  $\theta = \theta$ , a kind of *positive axis* for our polar coordinate system and  $R$  is the distance from the point  $P$  to the origin. Polar coordinates are defined in terms of ordinary Cartesian coordinates by computing and connecting the  $n$  points  $P_{ij}$ , for  $i = 1, \dots, n$ . It is calculated by using equation as follows:

$$P_{ij} = \begin{cases} x_{ij} = r \cos \theta_i \\ y_{ij} = r \sin \theta_i \end{cases} \quad (1)$$

where  $r \geq 0$   $0 \leq \theta < 2\pi$ , that every point  $P(x_{ij}, y_{ij})$  in the ordinary  $xy$ -plane (correlation plane) can be rewrite to  $(r, \theta)$ -answer which is, is a result of the fact of  $P$  lines on the circumference.

From these multiple relations, we called correlation of data on the coordinates of our point  $P$  satisfy the relation  $x_{ij}^2 + y_{ij}^2 = r_{ij}^2 (\cos^2 \theta_i + \sin^2 \theta_i) \rightarrow x_{ij}^2 + y_{ij}^2 = r_{ij}^2$  (so that, as we indicated, the point  $P(x_{ij}, y_{ij})$  and  $(\cos^2 \theta_i + \sin^2 \theta_i) = 1$ ) is on a circle of radius  $r$  centered at  $O$ ). So, we can find  $\theta$  by solving the equation as:

$$\tan \theta_i = \frac{y_{ij}}{x_{ij}} \rightarrow \theta_i = \arctan\left(\frac{y_{ij}}{x_{ij}}\right), \quad (2)$$

where  $\theta$  in the interval  $0 \leq \theta < 2\pi$ , let *arctangent* denoting the function by  $\arctan$  we see that:

$$\theta_i = \arctan\left(\frac{y_{ij}}{x_{ij}}\right) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } -\frac{\pi}{2} < \theta < \frac{\pi}{2}, \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{if } \frac{\pi}{2} < \theta \leq \frac{3\pi}{2}, \end{cases} \quad (3)$$

with the interpretation that  $\theta = \pm\pi/2$  corresponds to points on the real  $y$ -axis and  $\theta = 0$  corresponds to points on the real  $x$ -axis, that we called *correlation plot*.

### C. The distance between plots

We can use the theoretical *Pythagorus*<sup>1</sup> method to compute the distance between points in the plane in order to find the distance  $d$ . In Data mining we call *Centroid* [7] to calculate using equation as follows:

$$d_i = \frac{1}{n} \sum_{i=1}^n |c - v_i| \quad (4)$$

where  $C$  is the centroid or the correlation plot  $(x_{Coordinate}, y_{Coordinate})$ ,  $V_i$  is coordinates in the circumference  $(x_i, y_i)$ , and  $|C - V_i|$  is the distance between plots with the coordinates of  $i$  in the circumference, we see that:

$$|c - v_i| = \sqrt{(x_{coordinate} - x_i)^2 + (y_{coordinate} - y_i)^2} \quad (5)$$

### D. Dewey decimal classification

Dewey decimal classification was developed by Melvil Dewey in 1876. It is widely used in the library. Besides, there are many kinds of the books which unlimited of any field. That is the system used in more than 200,000 libraries in 135

countries. The Dewey decimal system divides the knowledge into 10 classes, in each class it is divided into 10 sub-classes and in each sub-class it is divided into 10 divisions accordingly. By using numbers as symbols with the purpose of easy to remember, it is popular to use with more than 30 languages translation around the world.

### E. Dewey decimal classification – Multiple relations

Dewey decimal classification – Multiple Relations or we call **DDC-MR**. It is a technical analysis classification multiple relations which was developed between Search engine and Dewey decimals classification. It focuses on the analysis of proportion in the content [3]. By using the library standard classification schemes, one keyword will be able to classify as deep as 4 levels which assigns number for notation [6], [7]. This scheme refers to DDC that does divide human knowledge into 10 classes in the first level, 100 subclasses in the second level, 1000 divisions in the third level and the last level or leaf node contains more than 10000 sections.

## III. DEFINITION TO CORRELATION

Our study of implementing methods is to study of correlation deformation connected by related radar graphs, and subsequently replaced by polar coordinates. One main concern of the study of implementing methods is to consider the shapes, quantities of content correlations, distances, correlation positions and directions of determined coordinates. Thus, in this research, the authors provide definitions for the purpose of comparing correlations before and after deformation and identifying advantages and implementing methods. For instance, a document pertaining to many sciences, when examined to find out whether it is a suitable representative of documents, has to be adapted so that the plot position is found and the plot of intrinsic correlation on the plane and boundary is consistent with that correlation. As such, a normal plane cannot explain correlations of that calculated position because the starting point originates from variables with multiple correlations. Below are definitions of keywords.

### A. Correlation plot

A correlation plot indicates a position of coordinates derived from a calculation of combined values of every correlation so that one position on the same area is identified. The point resulting from that calculation is titled in this research as a **correlation plot**, which is used to show or represent a position of each data set on the correlation plane referring to any correlation with  $n$  relevant contents. Correlations can be demonstrated in pairs  $(r, \theta)$ , where the first pair refers to only one plot and represents only one data set of distance and directional correlations of variables on polar coordinates. For example, one document containing a number of related contents is represented by  $n$  axis (with results shown in the form of a radar graph), and then calculated by mutual tension. Consequently, one plot in the form of  $(r, \theta)$  was acquired as seen in Figure 1.

<sup>1</sup>[http://en.wikipedia.org/wiki/Pythagorean\\_theorem](http://en.wikipedia.org/wiki/Pythagorean_theorem)

## B. Correlation plane

A correlation plane indicates the area where coordinates derived from calculated correlation points of data are located. The points require locations and addresses so normal planes cannot be applied in this research. The number of occurring correlations results from variables with multiple correlations. Therefore, the calculated values of pairs were not solely data derived from  $(x, y)$  axes, but also data resulted from tension among  $n$  axes that divided the plane within polar coordinates. In this research, the plane is called a **correlation plane** which is essential to distances and directional correlations especially loadings and depth directions. The intersection of  $n$  axis is called the origin and intersecting  $n$  axes divide the plane into  $n$  parts. Each part is called a **correlation boundary**.

## C. Correlation boundary

A correlation boundary indicates angle values from lines appearing on a correlation plane by determining the boundary of measurements of angles between  $x$  axis of the correlation plane and lines appearing on the plane. Boundaries are divided according to categories of applications. In this research, a **correlation boundary** is used to determine the correlation area and the content correlation level of each category. The area which is close to the center ( $O = \text{Origin}$ ) represents low density of the content of that category while the area which is far from the center represents high density of the content of that category or specificity highly consistent to that particular category. This is applicable for categorization of correlations with **DDC-MR** [3], [6]. For example, in order to divide the correlation boundary into 10 main scientific categories, each science has the width of  $36^\circ$  and the first correlation boundary starts from  $0^\circ$ . Then, a counterclockwise rotation was done in order to divide sessions and determine the correlation boundary of the subsequent categories starting at  $36^\circ$ ,  $72^\circ$ ,  $108^\circ$ ,  $144^\circ$ ,  $180^\circ$ ,  $216^\circ$ ,  $252^\circ$ ,  $288^\circ$  and  $324^\circ$ , respectively, as shown in Figure 2.

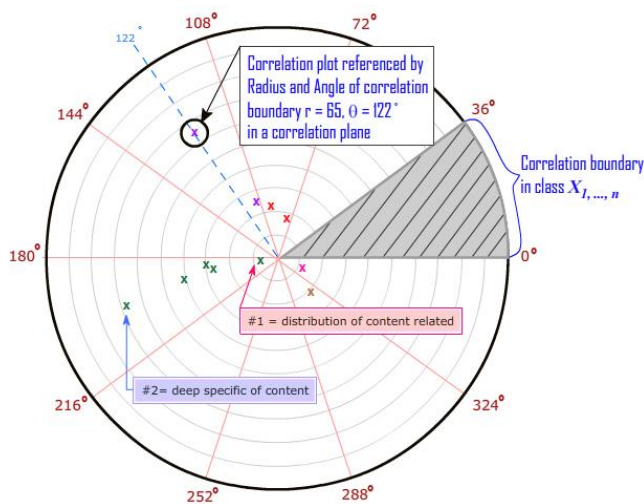


Figure 2. Example data for correlation plot refer to the distribution of content related and the deep specific of content in the correlation plane.

Figure 2 shows examples of determination of correlation plots on correlation planes, where correlation boundary in class  $X_1, \dots, n$  means the range of correlation boundary in each science. From the above example,  $n$  refers to 10 sciences, with the first science referring to a general class that has the correlation boundary of  $0^\circ - 35^\circ$ . The second science refers to a philosophy class with the correlation boundary of  $36^\circ - 71^\circ$ . The third science refers to a religion class with the correlation boundary of  $72^\circ - 107^\circ$  while the fourth science refers to a social sciences class with the correlation boundary of  $108^\circ - 143^\circ$ . The fifth science refers to a language class with the correlation boundary of  $144^\circ - 179^\circ$  and the sixth science refers to a pure science and mathematics class with the correlation boundary of  $180^\circ - 215^\circ$ . The seventh science refers to a technology and applied science class with the correlation boundary of  $216^\circ - 251^\circ$  while the eighth science refers to the arts and recreation class with the correlation boundary of  $252^\circ - 287^\circ$ . The next science is a literature class with the correlation boundary of  $288^\circ - 323^\circ$ . And the last science is a history and geography class with the correlation boundary of  $324^\circ - 360^\circ$ , respectively by Dewey decimal classification (DDC).

Positions of occurring points, or correlation plots, can be employed to refer to variables with  $n$  correlations. Each correlation differs in quantity and direction leading to different distances between coordinates on the correlation plane and the origin. Therefore, in accordance with the DDC classification of books, a widely practiced technique among libraries, if each plot is replaced by a set of books, the calculated correlation plot will be replaced by related contents of books, and the correlation plane will be replaced by areas of correlation of scientific content structure respectively. Dense plots are lines appearing in the direction with correlations within the correlation boundary. The plot which is very far from the center means that a book containing very specific and in-depth contents of that science. Since force loading and directions of variables are highly related and the plot which is very close to the center also means that the book is specific to that science, but does not have contents related to many sciences, as seen in Figure 2 (#1 and #2), if the loading and direction in each science are highly related in terms of proportion, that book will have contents related to many sciences. Additionally, redundant plots will bring about a different pattern of correlations of books with related contents. We, then, realize which books have the same content and what kind of content they have. It is possible to state that **correlation plots**, **correlation planes** and **correlation boundaries** have continuously related meanings and are major basic elements of application of multiple correlations.

## IV. CONCEPT TO APPLICATION

### A. Conceptual

The concept of calculating variables that have correlation is an analyzing technique developed from a mathematic method: a proper representative of data is identified by calculating total tension values and presenting them in the

form of polar coordinates of correlation. The objectives of this technique are to demonstrate similarity of data in the same area and explicitly explain levels of relations that, the process as follows in Figure 3.

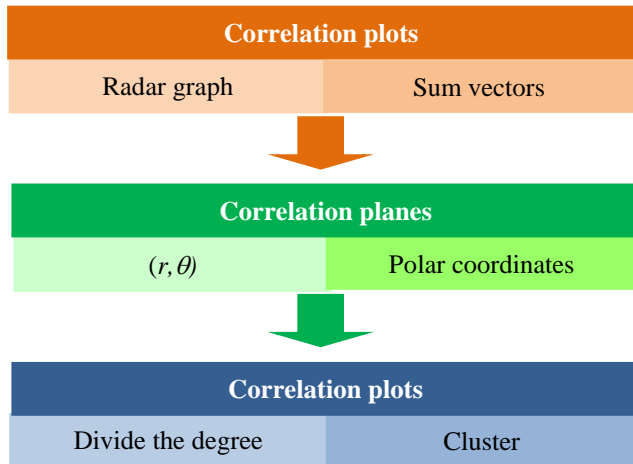


Figure 3. The concept of Correlation plots, Correlation plane and Correlation boundary.

- Changing all relations of variables to correlation plots is a process of summing vectors, where all classified correlations that can be clearly seen on a radar graph of one document are calculated so that one plot with the pair value of  $(r, \theta)$  is acquired and represents all relations of that document.

- Locating the position of a document with correlation plane, as seen in the above process, yields a pair value of  $(r, \theta)$  that represents the document. The pair value is then plotted, using the principle of polar coordinate determining the plane and  $(x, y)$  axes instead of applying its value only. Therefore, if we want to present several documents simultaneously, we have to have a number of axes to indicate the position of each document and determine a correlation plane so that all documents can be at their  $(r, \theta)$  values on the determined plane in that particular area. These way data sets are overlapped and not presented one by one. As such, no matter how the  $(r, \theta)$  value is calculated, that document will always be on that axis.

- Identifying the boundary section of an area is a process of grouping correlation planes used to indicate the position of each document and overlapping a number of polar coordinates so that several unseen axes are produced. Therefore, to categorize that data sets or document in a clear manner, the correlation boundary of those axes needs to be determined in accordance with the number of sections of the sample group

#### B. Analysis

If we use this method to analyze and categorize data, as seen in Section 3 with examples, and use the boundary on DDC-MR to categorize the data on the correlation plane, the document analyzed by the DDC-MR process will be able to locate that position. The correlation boundary will be 10 groups in the DDC main section. However, if we apply this concept and plot the boundary by competency, the boundary

of the data will be changed along with the number of sections of competency. If the size of competency sections is 10 of DDC-MR, then the same size of boundary can be applied, but if the sections decrease or increase in size, the degree size of the applied boundary will change accordingly.

## V. EXPERIMENTS AND RESULTS

This section performance of correlation plot, correlation plane and correlation boundary are shown three ways. The first way is *complexity of correlation*. It is used to explain the multi variances with multiple relations; if these correlations are high performance they should represent the different data in that correlation plane. The second way is *accuracy of classifying and analyzing* with the different multiple relations; we test correctness by articles, documents library and competency data. And the last way is *features to use these correlations classification*.

### A. Data sets

In our experiments, we used a collection of multiple relations of data from three examples given below.

- *Academic articles*: This data from a national conference disciplines in the computer and information technology which were published during 2005 to 2010, and we provided the dataset used 3 sections: Title, Abstract and Keyword. This data has multiple relations by DDC-MR in level 3 of DDC to classify 1,000 classes. The total number of articles is 700.

- *Documents library*: This data from the document library in the multidisciplinary amounting to 100 documents and we provided the data set used 3 sections: Title, Table of content and Index. Each document contains multiple relations links to other content which are related to the document.

- *Competency data*: This data from evaluate 10 out of 18 principles of competencies evaluation Spencer [10], [11], to select personnel basic competencies. There are: Achievement orientation, Analytical thinking, Conceptual thinking, Customer service orientation, Developing others, Impact and influence, Information seeking, Teamwork and cooperation, Team leadership and Self-confidence.

### B. Experiments

We used the correlation plot, correlation plane and correlation boundary provided by the multiple relations of multi variances to computed our experiments. The correlation plot is a coordinates from the computed of all relationships, the correlation plane is an area coordinates arising from correlation plot and the correlation boundary is a range between the degrees of set. In this experiment, we applied the radar graph provided under the correlation and set the number of academic articles to 1,000 classes, set the documents library to 100 classes and set the number of competency data to 10 classes. For text classification process of academic articles and documents library, we used DDC-MR and competency data from analytical to perform the experiments.



### C. Evaluation Metrics

The standard performance metrics for evaluation the classification used in the experiments is accuracy. These metrics assume the prediction process evaluation based on the counts of test records correctly and incorrectly predicted. That shows the confusion matrix for a binary classification problem. Each entry  $f_{ij}$  in the confusion matrix denotes the number of records from class  $i$  predicted to be of class  $j$ , which is defined as follows:

$$\text{Accuracy} = \frac{\text{the number of correctly classified test documents}}{\text{total number of test documents}} \quad (6)$$

where Accuracy represents the percentage of correct predictions in total predictions. We used Accuracy as our choice of evaluation metric to report prediction experiments because it can be expressed in terms and most classification seek models that attain the highest accuracy when applied to the test set. In our previous experiments we have seen that Accuracy provide the experimental schemes in terms of prediction performance.

### D. Experimental results

The innovation of our paper is presented how to analyze multi variants data by plot in the correlation plane. In this research, the experimental results of document relevant focus on three issues such as: accuracy to classification, time in process and the features interesting.

- **Accuracy:** The experimental results on accuracy classification by comparing between Correlation plot, K-Mean clustering, Hierarchical clustering and Factor analysis. We found out that, the correlation plot method showed that the accuracy is higher than other methods. As shown in Table 1.

TABLE I. COMPARISON THE ACCURACY OF CLASSIFICATION RESULTS IN EACH CLASS.

Model	Accuracy (%)			
	Correlation plot	K-Mean	Hierarchy	Factor Analysis
C1	97.66	95.31	97.66	92.97
C2	93.75	97.66	97.66	93.75
C3	93.75	90.63	86.72	87.50
C4	92.97	92.97	92.19	86.72
C5	92.19	78.91	64.06	59.38
C6	86.72	79.69	67.19	72.88
C7	88.50	88.28	85.94	70.31
C8	93.75	85.16	93.75	92.19
C9	97.66	95.31	91.41	92.97
C10	98.44	97.66	96.86	97.66

Table 1 shows experimental results comparing the accuracy of 10 clusters describing how the Correlation plot method had the best accuracy. Thus, if we compared with other methods, almost every parameter has a very high

accuracy such as: C1 = 97.66%, C2 = 93.75%, C3 = 93.75%, C4 = 92.97%, C5 = 92.19%, C6 = 86.72%, C7 = 88.50%, C8 = 93.75%, C9 = 97.66 and C10 = 98.44%, which we considering all clusters. The accuracy of correlation plot is close to K-Mean clustering. It means both of method can be used to classify data in this research. If we compared in all clusters, we find out that K-Mean clustering had problem in 2 clusters; that were C5 and C6, The accuracy was less than 80% while the accuracy correlation plot method were more than 80% in all clusters. Furthermore, the accuracy of Hierarchical clustering and Factor analyses were similar and while accuracy of some clusters was less than 80%. This means the effectiveness of cluster is low, as shown in Figure 4.

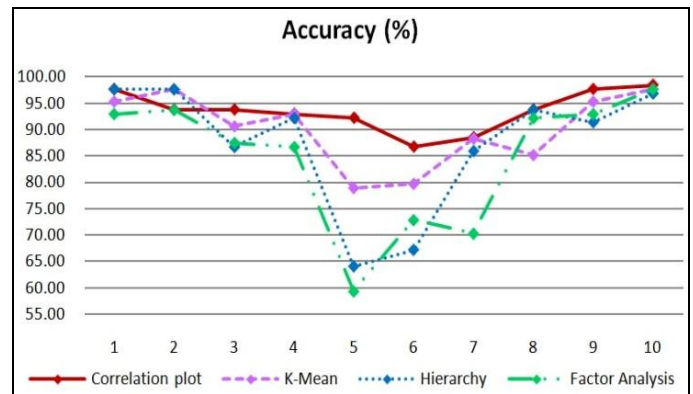


Figure 4. Comparison the accuracy of four classifications.

- **Time:** The experimental results of time in process shows that, if we applied to the Polar coordinate, the correlation plot methods had best effective use when comparing to other methods. As shown in Table 2.

TABLE II. TIME COMPARISON RESULTS

Model	Time (Second)			
	Correlation plot	K-Mean	Hierarchy	Factor Analysis
100	0.0078	0.0262	0.1094	0.0469
200	0.0352	0.0625	0.168	0.0859
300	0.0859	0.1055	0.1445	0.1953
400	0.0664	0.1016	0.1953	0.2500
500	0.0977	0.1133	0.1836	0.4258
600	0.1029	0.1159	0.2013	0.4369
700	0.1033	0.1309	0.2145	0.5012

Table 2 shows experimental results comparing the time in processing with 4 methods. We test with difference 7 data sets size of data. The first data sets had 100 documents, the second had 200 documents, the third had 300 documents, the fourth had 400 documents, the fifth had 500 documents, the sixth had 600 documents and the last data sets had 700 documents. The results described correlation plot used the less time or minimum in processing. Thus, if we analysis with



the 7 data sets using similar or nearest time in processing, the increasing of the data amount does not effect to time in processing because this method does not need to recalculate every time whenever adding new information. Thus, original data will remain the same position and same cluster. Factor analysis shows the results effect with time in processing, if we change the amount of data it will spend more time, as shown in Figure 5.

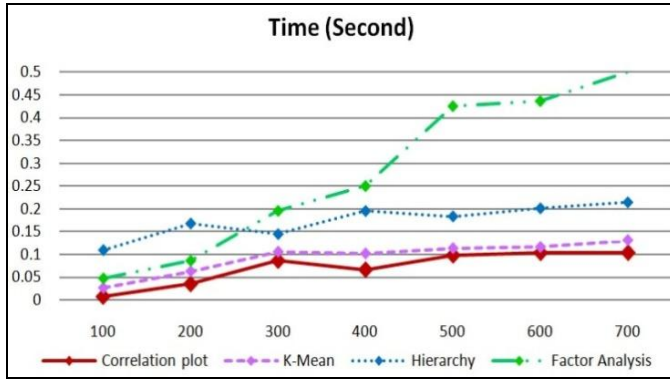


Figure 5. Time comparison of four classifications.

Table 3 shows experimental results comparing accuracy and time in processing between Correlation plot with K-Mean clustering, Hierarchical clustering and Factor analysis. This is to test performance cluster with 700 articles in the first time and we repeated the second time with 100 documents library. The results shown that, correlation plot were the most accuracy in academic article which was 93.54% while document library was 90.33%. The process in minimum for academic article was 0.1033 and document library was 0.0452. In this research, the value of correlation plot was similar with K-Mean clustering in accuracy and time in processing, this means that the two methods can be used to classification in this data sets. Furthermore, Factor analysis had the lowest accuracy in academic article which was 84.63% and document library which was 78.54%. They are lower than the statistical acceptance and lower with the criteria by the researcher which the accuracy must be greater than 80%. In addition, it used the most of time in processing academic article which was 0.5012 and document library which was 0.1959. They are over than the time criteria that must be less than 15 seconds.

TABLE III. EVALUATION RESULTS ON CLASSIFICATION

Model	Accuracy (%)		Time (second)	
	Article (700)	Doc. (100)	Article (700)	Doc. (100)
Correlation plot	93.54	90.33	0.1033	0.0452
K-Mean	90.16	88.63	0.1309	0.0959
Hierarchy	87.34	85.21	0.2145	0.1594
Factor Analysis	84.63	78.54	0.5012	0.1959

TABLE IV. FEATURES COMPARISON RESULTS ON CLASSIFICATION.

Features	Model			
	Correlation plot	K-Mean	Hierarchy	Factor analysis
1. Easy to understand	/	/	/	/
2. Segmentation is clear	/	/	/	/
3. To break the color of the group	/	/	/	x
4. Display hierarchical data	/	x	/	x
5. Display depth of information	/	x	x	x
6. Display specific information on each group	/	x	x	x
7. Display the direction and distance	/	x	/	/
8. Display data in multiple groups simultaneously	/	x	x	x
9. The ability to compare data	/	x	x	x
10. Do not adjust the scale display	/	x	x	x
11. Do constants in the group	/	x	x	x
12. The amount of data does not affect to the process	/	/	x	x
13. Time in process < 15 second*	/	/	x	x
14. Accuracy > 85% *	/	/	/	x

\* From evaluation results in Table 3

• *Features:* From the experimental results and analysis with the performance of classification shows that, if we apply data with the Polar coordinate it will increase ability of classification and will get more features interesting. From table 4, we used the same data set-compared with the elements and dominant features of correlation plot method effected to more classification interesting which includes 14 features as follows: 1) *Easy to understand* when we used it in the first time. 2) *Segmentation is clear* because we had the line segment. 3) *To break the color of the group* for more separated data. 4) *Display hierarchical of data* in each cluster because it had radius of the circle line to compute. 5) *Display depth of information* to relevant in each cluster and could refer to the distribution of content related. 6) *Display specific information* on each group, if we plot near the center, it means that the content is similar with the other cluster but, if we plot far from the center, it means that the content is more specific in the cluster. 7) *Display the direction and distance* in each cluster, if we know that we can predict the road map and fulfill the knowledge in each cluster. 8) *Display data in multiple groups simultaneously*. It means that the ability to display more than in one cluster in the same time such as: 4 clusters or 10 clusters depending on the desired number of clusters. 9) *The ability to compare data* with pie charts is useful for comparing the contributions of different clusters to the total value of a variable. We can also compare two identical nominal data sets under different conditions, especially at different time. 10) *Do not adjust the scale display* when we show all information because in some visualization [7] we need for larger visualize to clearly information. 11) *Do constants in the group* mean that if we added new data, the original data plot is in the same cluster and does not compute to new cluster. This method was

difference with other methods because every parameter had affected to the classification. 12) *The amount of data does not affect to the process.* From the results in Figure 5 it can be proved that the number of data does not a problem for this method. 13) *Time in process < 15 second\**, and 14) *Accuracy > 85% \** from the evaluation results in Table 3 shows the performance of the experimental.

## VI. CONCLUSION

In an attempt to improve the performance of correlation plot, correlation plane and correlation boundary, we propose an innovative method of multiple correlation application, a method of polar coordinate application so that data are categorized in an effective manner regardless of their type and quantity in the sample group. With correlation plot, correlation plane and correlation boundary, we can present data that are representatives of that relation: they are applied to identify relations of data and categorize data in many aspects. According to a test, it is found out that if we use boundary on DDC-MR to categorize a data sets, it could be categorized into 10 groups of the DDC main section. However, if this concept is applied to other data sets, the degree of boundary will be changed along with the number of sections of those data sets. The authors believe that correlation application can give practical explanation on analysis and data presentation to users as follows.

1) *Advantages to users:* correlation application is clear, comprehensive and reliable. It is a data presentation method in the form of bar graphs with explicit lines dividing boundaries of each group. Different colors can be applied to different bars for the aesthetic and distinct purposes. With correlation application, problems of various levels of users' basic knowledge can be reduced because it is a familiar presentation method widely used in data analysis explanation.

2) *Advantages of knowledge management in organizations* with an emphasis on enhancement of knowledge storage capacity: most organizations focus on storing, collecting, exchanging, transferring and publishing data and do not consider how those data are related. Hence, in order to optimize organizations' benefits, relations of those data should be identified. This correlation method can help analyze and present different levels of data with related contents so knowledge. It can be implemented in a useful way. If we can effectively explain data, those data will support our work and foster ongoing changes.

3) *Advantages of integration:* different kinds of information that is related or has the same content are applied and transformed into new knowledge that is easy for users to understand and implement. The authors believe that correlation application can help synthesize, analyze and explain connections of information contents. For example, relations of different courses can be identified by correlation application: courses that are close together, overlapped or missing can be replaced by one another because of their high similarity, or courses can be combined into one new course relevant to former basic courses.

4) *Advantages of data mining:* previous research placed emphasis on sorting data or categorizing relevant data into the same group and paid no attention to relations of data contents. Correlation application, apart from being able to precisely and rapidly sort data (referring to test results in Table 3), can explain appearing relations of information at the content level and classify levels of relations of information within a group in a clear manner.

## REFERENCES

- [1] "Data Analysis: Displaying Data – Graphs – 1", Texas State Auditor's Office, Methodology Manual, rev. 5/95, [http://www.preciousheart.net/chaplaincy/Auditor\\_Manual/11graphd.pdf](http://www.preciousheart.net/chaplaincy/Auditor_Manual/11graphd.pdf)
- [2] S. T. Yen, "Using Radar chart to display Clinical data".
- [3] W. Lertmahakiat and A. Mingkhwan, "Information Retrieval by Multiple DDC Relational Classification," in Proceeding on NCCIT'08, the 4<sup>th</sup> National Conference on Computing and Information Technology, Thailand. 22 – 23 May 2008.
- [4] J. Wang, J. Li and W. Hong, "Feature Extraction and Classification for Graphical Representations of Data," Springer-Verlag Berlin Heidelberg, vol. 5226, pp. 506 – 513, 2008.
- [5] J. Wang, W. Hong and X. Li, "The New Graphical Features of Star Plot for K Nearest Neighbor Classifier," Springer-Verlag Berlin Heidelberg, vol. 4682/2007, pp. 926 – 933, July 2007.
- [6] J. Watthananon and A. Mingkhwan, "Multiple Relation Knowledge Mapping Using Rectangular and Polar Coordinates," in Proceeding on PGNET'09, the 10th Annual Postgraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting, Liverpool John Moores University, UK. 21 – 22 June 2009.
- [7] J. Watthananon and A. Mingkhwan, "A Magnify Classification technique for group of Knowledge Using DDC-MR," in Proceeding on NCCIT'09, the 5<sup>th</sup> National Conference on Computing and Information Technology, Thailand. 22 – 23 May 2009.
- [8] A. Klippel, F. Hardisty, R. Li and C. Weaver, "Color Enhanced Star Plot Glyphs Can Salient Shape Characteristics be Overcome?," *Journal of Cartographica*, 44:3, pp. 217 - 231, 2009.
- [9] J. Peng, W. Yang and Y. Li, "Shape Classification Based on Histogram Representation in Curvature Scale Space," *IEEE Transaction on Computational Intelligence and Security*, vol. 2, pp. 1722 – 1725, 2006.
- [10] L. M., Spencer, S. M., Spencer. "Competence at Work – Models for Superior Performance," Wiley, New York, NY, 1993.
- [11] P. Chintanaporn and A. Mingkhwan, "A Study of Information Retrieval Algorithm for Multiple Relation Competency Team Recruitment," in Proceeding on PGNET'09, the 10th Annual Postgraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting, Liverpool John Moores University, UK. 21 – 22 June 2009.
- [12] N. Sukumar and A. Tabarraei, "Polygonal Interpolants: Construction and Adaptive Computations on Quadtree Meshes," in Proceeding on ECCOMAS 2004, July 2004, pp. 1 – 9.
- [13] K. Mohseni and T. Colonius, "Numerical Treatment of Polar Coordinate Singularities," *Journal of Computational Physics*, vol. 157, pp. 787 – 795, 2000.
- [14] J. Demsar, G. Leban and B. Zupan, "FreeViz-An intelligent multivariate visualization approach to explorative analysis of biomedical data," *Journal of Biomedical Informatics*, vol. 40, pp. 661 – 671, 2007.
- [15] M. Meyer, H. Lee, A. Barr and M. Desbrun, "Generalized Barycentric Coordinates on Irregular Polygons," *Journal of Graphics Tools*, vol. 7, pp. 13 – 22, November 2002.
- [16] A. Klippel, F. Hardisty, and C. Weaver, "Star Plots: How Shape Characteristics Influence Classification Tasks". *Cartography and Geographic Information Science*, 36:2, pp. 149 - 163, 2009.
- [17] S. N. Patnaik and D. A. Hopkins, "Completed Beltrami-Michell formulation in Polar coordinates," *International Journal of Physical Sciences*, vol. 2 (5), pp. 128 – 139, May 2007.

AUTHORS PROFILE



**Julaluk Watthananon** received B.Sc. degree in Information System from Rajamangala University of Technology Thanyaburi, M.S. degree in Information Technology from King Mongkut's University of Technology North Bangkok. Currently, she is a Ph.D. candidate in the Faculty of Information Technology at King Mongkut's University of Technology North Bangkok. Her current research interests Knowledge Management, Data Mining and Information Retrieval.



**Sageemas Na Wichian** received Ph. D. in Educational Research Methodology at Chulalongkorn University. She has been an educator for over ten years, teaching in universities in the areas of diversity, Psychology and Research for Information Technology. She is currently working as a lecturer at King Mongkut's University of Technology North Bangkok, Thailand. Her research focuses is on advanced research methodology, Industrial and Organizational psychology.



**Anirach Mingkhwan** received B.Sc. degree in Computer Science from King Mongkut's University of Technology North Bangkok, M.Sc. degree in Computer and Information Technology from King Mongkut's Institute of Technology Ladkrabang and Ph.D. degree in Computer Network from Liverpool John Moores University. His current research interests Computer Network and Information Security, Wireless and Mobile Ad Hoc Networks, Knowledge management and Ontology. Currently, he is an Assistant Professor in the Faculty of Industrial and Technology Management at King Mongkut's University of Technology North Bangkok, Thailand.

# Structural Analysis of Bangla Sentences of Different Tenses for Automatic Bangla Machine Translator

Md. Musfique Anwar, Nasrin Sultana Shume and Md. Al-Amin Bhuiyan

Dept. of Computer Science & Engineering, Jahangirnagar University, Dhaka, Bangladesh

Email: musfique.anwar@gmail.com, shume\_sultana@yahoo.com, alamin\_bhuiyan@yahoo.com

## Abstract

*This paper addresses about structural mappings of Bangla sentences of different tenses for machine translation (MT). Machine translation requires analysis, transfer and generation steps to produce target language output from a source language input. Structural representation of Bangla sentences encodes the information of Bangla sentences and a transfer module has been designed that can generate English sentences using Context Free Grammar (CFG). The MT system generates parse tree according to the parse rules and a lexicon provides the properties of the word and its meaning in the target language. The MT system can be extendable to paragraph translation.*

## Keywords:

*Machine Translation, Structural representation, Context Free Grammar, Parse tree, Lexicon etc.*

## 1. Introduction

Machine translator refers to computerized system responsible for the production of translation from one natural language to another, with or without human assistance. It excludes computer-based translation tools, which support translators by providing access to on-line dictionaries, remote terminology databanks, transmission and reception of texts, etc. The core of MT itself is the automation of the full translation process. Machine translation (MT) means translation using computers.

We need to determine a sentence structure at first using grammatical rules to interpret any language. Parsing or, more formally, syntactic analysis, is the process of analyzing a text, made of a sequence of tokens (for example, words), to determine its grammatical structure with respect to a given formal grammar. Parsing a sentence produces structural representation (SR) or parse tree of the sentence [1].

Analysis and generation are two major phases of machine translation. There are two main techniques concerned in analysis phase and these are morphological analysis and syntactic analysis.

Morphological parsing strategy decomposes a word into morphemes given lexicon list, proper lexicon order and different spelling change rules [2]. That means, it incorporates the rules by which the words are analyzed. For example, in the sentence - "The young girl's behavior was *unladylike*", the word

"Unladylike" can be divided into the morphemes as Un – not (prefix), Lady – well behaved female (root word), Like – having the characteristics of (suffix). Morphological information of words are stored together with syntactic and semantic information of the words.

The purpose of syntactic analysis is to determine the structure of the input text. This structure consists of a hierarchy of phrases, the smallest of which are the basic symbols and the largest of which is the sentence. It can be described by a tree known as parse/syntax tree with one node for each phrase. Basic symbols are represented by leaf nodes and other phrases by interior nodes. The root of the tree represents the sentence.

Syntactic analysis aims to identify the sequence of grammatical elements e.g. article, verb, preposition, etc or of functional elements e.g. subject, predicate, the grouping of grammatical elements e.g. nominal phrases consisting of nouns, articles, adjectives and other modifiers and the recognition of dependency relations i.e. hierarchical relations. If we can identify the syntactic constituents of sentences, it will be easier for us to obtain the structural representation of the sentence [3].

Most grammar rule formalisms are based on the idea of phrase structure – that strings are composed of substrings called phrases, which come in different categories. There are three types of phrases in Bangla- Noun phrase, Adjective Phrase and Verb Phrase. Simple sentences are composed of these phrases. Complex and compound sentences are composed of simple sentences [4].

Within the early standard transformational models it is assumed that basic phrase markers are generated by phrase structure rules (PS rules) of the following sort [5]:

$S \rightarrow NP \text{ AUX } VP$   
 $NP \rightarrow ART \text{ N}$   
 $VP \rightarrow V \text{ NP}$

PS rules given above tell us that a S (sentence) can consist of, or can be expanded as, the sequence NP (noun phrase) AUX (auxiliary verb) VP (verb phrase). The rules also indicate that NP can be expanded as ART N and that VP can be expressed as V NP.

This paper implements a technique to perform structural analysis of Bangla sentences of different tenses using Context Free Grammar rules.

## 2. Bangla Sentences Structure

In Bangla language, a simple sentence is formed by an independent clause or principal clause. Example: শাহেদ কুলে যায়

A complex sentence consists of one or more subordinate clause within a principle clause [2]. As for example, যদি সবুজ বাতি জ্বলে তাহলে রাস্তা পারাপার করে। Bangla compound sentence is formed by two or more principal clauses joined by an indeclinable/conjunctive (অব্যয় পদ) Example:

অধিভিবন্দ সদায় উপস্থিত হন এবং সভাপতি কার্যক্রম শুরু করেন

Types of Bangla tense are given below in Fig. 1:

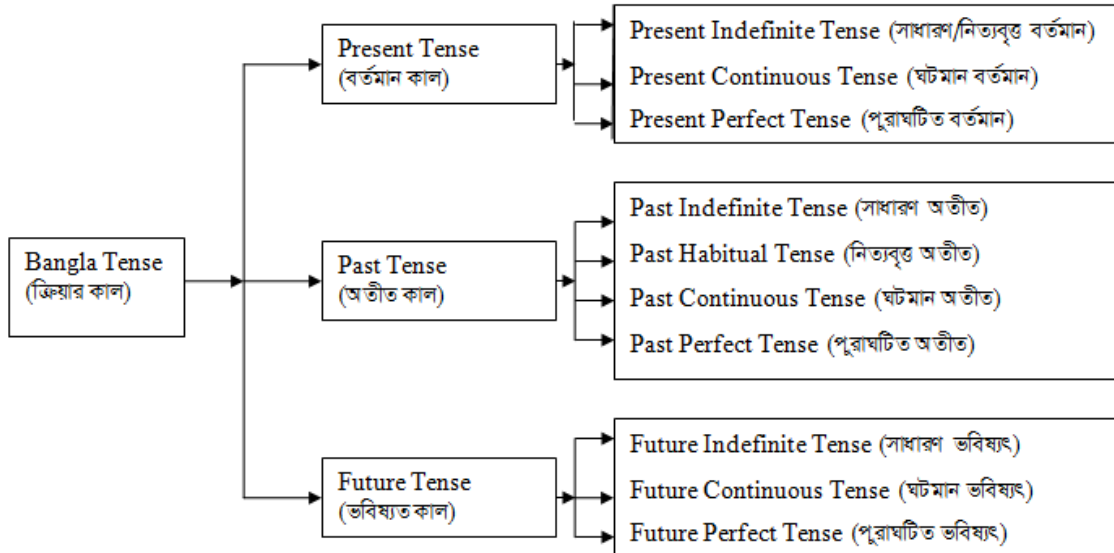


Fig. 1 Types of Bangla tense

### 2.1 Basic Structural Difference between Bangla and English Language

Following are the structural differences between Bangla and English languages:

- The basic sentence pattern in English is subject + verb + object (SVO), whereas in Bangla it is - subject + object + verb (SOV). Example:

English: I (S) eat (V) rice (O)

Bangla: আমি (S) ভাত (O) খাই (V)

- Auxiliary verb is absent in Bangla language.

Example: I (Pronoun) am (Auxiliary verb) reading (Main verb) a (Article) book (Noun)

Bangla: আমি (Pronoun) একটি (Article) বই (Noun) পড়িতেছি (Main verb)

- Preposition is a word placed before a noun or pronoun or a noun-equivalent to show its relation to any other word of the sentence [6]. In Bangla language, bivakti will place after noun or pronoun or a noun-equivalent. Example:

English: The man sat on the chair

Bangla: লোকটি চেয়ারটিতে বসল, here 'তে' is bivakti

### 2.2 Structural Transfer from Bangla to English

Parsing is the process of building a parse tree for an input string. We can extract the syntactic structure of a Bangla sentence using any of the two approaches: i) top-down parsing ii) bottom-up parsing.

#### 2.2.1 Top-Down Parsing

Top-down parsing starts at the most abstract level (the level of sentences) and work down to the most concrete level (the level of words). An input sentence is derived using the context-free grammar rules by matching the terminals of the sentence. So, given an input string, we start out by assuming that it is a sentence, and then try to prove that it really is one by using the grammar rules left-to-right. That works as follows: If we want to prove that the input is of category  $S$  and we have the rule  $S \rightarrow NP VP$ , then we will try next to prove that the input string consists of a noun phrase followed by a verb phrase.

#### 2.2.2 Bottom-Up Parsing

The basic idea of bottom up parsing is to begin with the concrete data provided by the input string --- that is, the words we have to parse/recognize --- and try to build more abstract high-level information.

Example: Consider the Bangla sentence “বালকটি চা পান করে”. To perform bottom-up parsing of the sentence using the following rules of the context-free grammar,

<SENTENCE>  $\rightarrow$  <NOUN-PHRASE> <VERB-PHRASE\*>

<NOUN-PHRASE>  $\rightarrow$  <CMPLX-NOUN> |

<CMPLX-NOUN> <PREP-PHRASE\*> | <ART>

<ADJ> <NOUN> <PREP-PHRASE\*>



<VERB-PHRASE> → <CMPLX-VERB> |  
 <CMPLX-VERB> <PREP-PHRASE\*> | <CMPLX-  
 VERB> <PREP-PHRASE\*>  
 <PREP-PHRASE> → <PREP> <CMPLX-NOUN\*> |  
 <PREP> <PRONOUN>  
 <CMPLX-NOUN> → <ARTICLE> <NOUN> |  
 <NOUN> | <PRONOUN>  
 | <NOUN> <PRONOUN> <NOUN>  
 <CMPLX-VERB> → < MAIN-VERB> | < MAIN-  
 VERB> <NOUN-PHRASE\*>

During the bottom-up parsing of the Bangla sentence “বালকটি চা পান করে”, we obtain the syntactical grammatical structure NOUN ARTICLE NOUN MAIN-VERB.

The syntactic categories in the resulting grammatical

structure are then replaced by the constituents of the same or smaller unit till a SENTENCE is obtained, which is shown below:

Input    বালকটি চা পান করে  
 Sentence    NOUN ARTICLE NOUN MAIN-VERB  
               → NOUN-PHRASE NOUN-PHRASE  
                   MAIN-VERB  
               → NOUN-PHRASE CMPLX-VERB  
               → NOUN-PHRASE VERB-PHRASE  
               → SENTENCE

### 3. Proposed MT Model

The model proposed model for structural analysis of Bangla sentences is shown in Fig. 2.

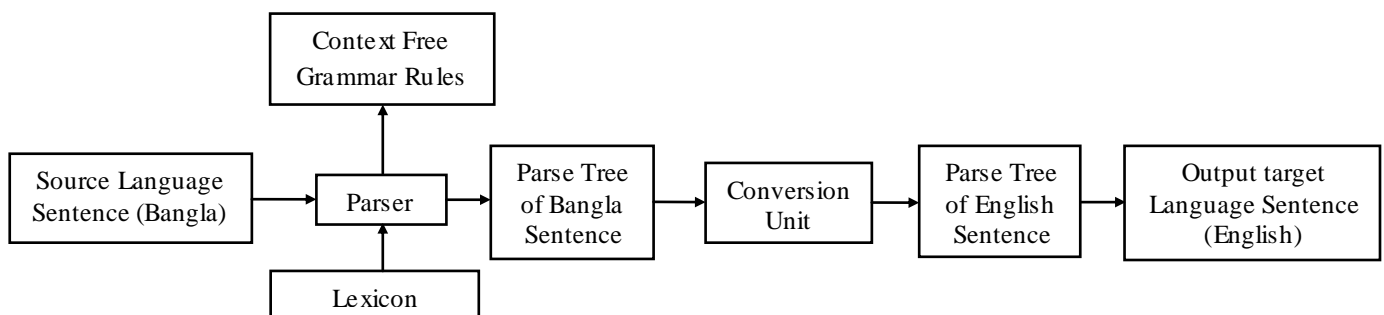


Fig. 2 Block diagram of proposed MT model

#### 3.1 Description of the Proposed Model

The proposed MT system will take a Bangla natural sentence as input for parsing. Stream of characters are sequentially scanned and grouped into tokens according to lexicon. The words having a collective meaning are grouped together in a lexicon. The output of the Tokenizer of the input sentence “Cheleti Boi Porche” is as follows [1] [4]:

TOKEN = (“Chele”, “Ti”, “Boi”, “Por”, “Che”).

The parser involves grouping of tokens into grammatical phrases that are used to synthesize the output. Usually, the phrases are represented by a parse tree that depicts the syntactic structure of the input.

A lexicon can be defined as a dictionary of words where each word contains some syntactic, semantic, and possibly some pragmatic information. The entries in a lexicon could be grouped and given by word category (nouns, verbs, prepositions and so on), and all words contained within the lexicon listed within the categories to which they belong [1] [4] [5] [7]. In our project, the lexicon contains the English meaning and parts of speech of a Bangla word.

A context-free grammar (CFG) is a set of recursive rewriting rules (or productions) used to generate

patterns of strings. It provides a simple and precise mechanism for describing the methods by which phrases in some natural language are built from smaller blocks, capturing the “block structure” of sentences in a natural way. Such as noun, verb, and preposition and their respective phrases lead to a natural recursion because noun phrase may appear inside a verb phrase and vice versa. The most common way to represent grammar is as a set of production rules which says how the parts of speech can put together to make grammatical, or “well-formed” sentences [8].

In the conversion unit, an input sentence is analyzed and a source language (SL) parse tree is produced using bottom-up parsing methodology. Then the corresponding parse tree of target language (TL) is produced. Each Bangla word of the input sentence is replaced with the corresponding English word from the lexicon in the target (English) parse tree to produce the target (English) language sentence.

Structural Representation (SR) is a process of finding a parse tree for a given input string. For example, the parse tree of the input sentence “বালকটি চা পান করে” and the corresponding parse tree of the English sentence “The boy drinks tea” is shown in Fig. 3.

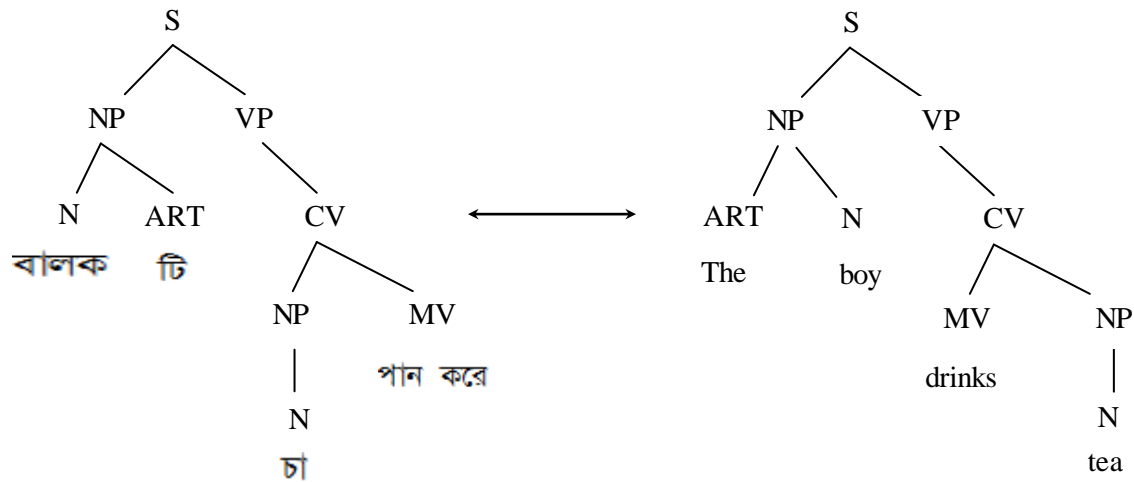


Fig. 3 Bangla and English Parse tree of the sentence “বালকটি চা পান করে”

#### 4. Implementation of the Proposed Model

Flow-chart of the proposed MT model is given below:

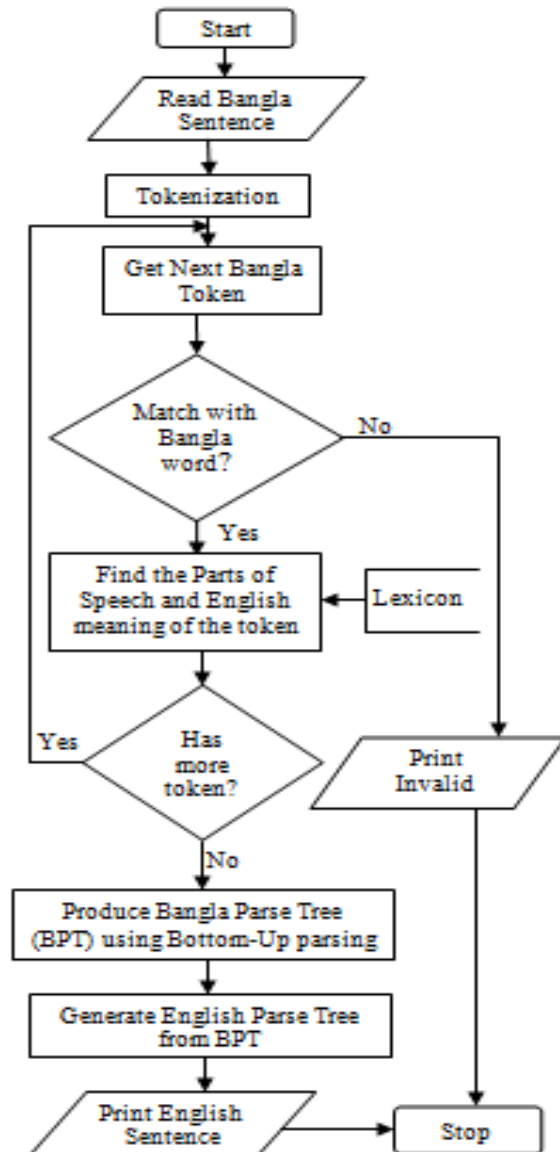


Fig. 4 Flow-chart of the proposed MT Model

After executing the above procedure according to the Flow-chart, it is possible to translate a Bangla sentence into corresponding English sentence.

#### 5. Experimental Results

Several experiments were conducted to justify the effectiveness of the proposed MT model. Success rate for different types of sentences is shown in Fig. 5. Fig. 6 illustrates the snapshot of the implemented method.

Table 1: Success rate for different types of sentences

Type of Sentences	Total no. of sentences	Correctly performed machine translation	Success rate (%)
Simple	770	745	96.75
Complex	540	517	95.74
Compound	360	338	93.89

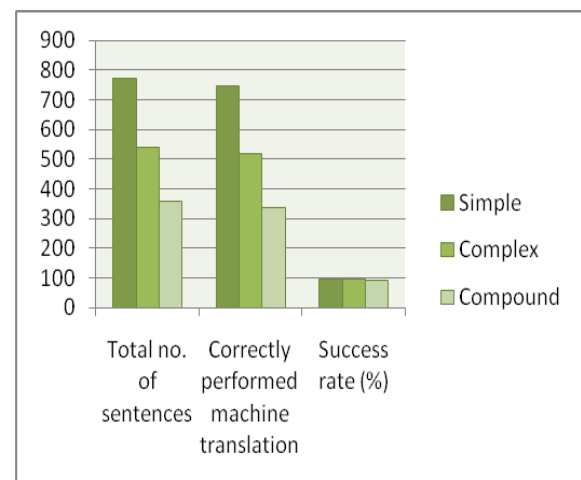


Fig. 5 Success rate for different types of sentences

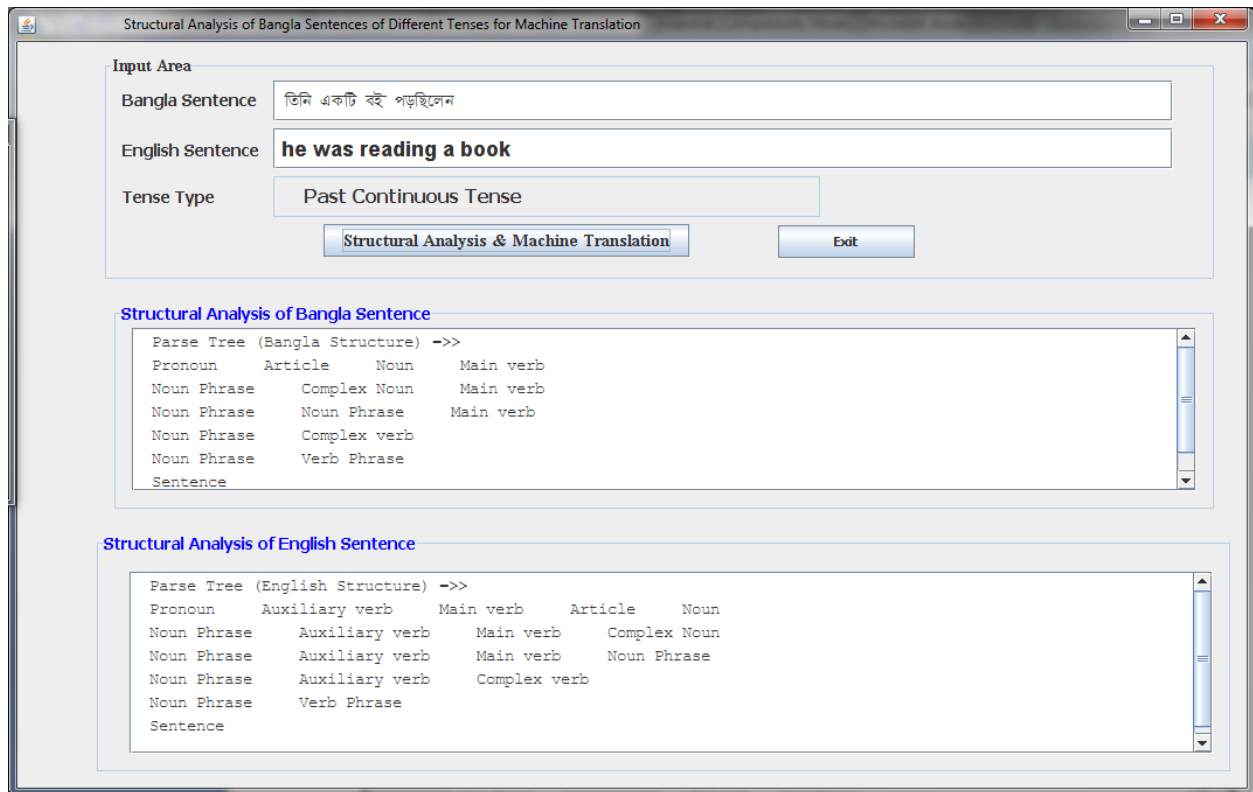


Fig. 6: Sample output of the program for the complex sentence

## 6. Conclusion

This paper mainly focuses on the structural analysis phase of how to build parse tree of a given Bangla sentence according to CFG. The translation process is then applied to the Source Language (SL) Tree to obtain a tree with target language words (TL Tree). Finally, the output sentence in the target language is extracted from this tree in the target language and also indicates the type of the tense. But the sentences composed of idioms and phrases are beyond the scope of this project.

## References

- [1] M. M. Hoque and M. M. Ali, "A Parsing Methodology for Bangla Natural Language Sentences", Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 277-282 (2003).
- [2] S. Dasgupta and M. Khan, "Feature Unification for Morphological Parsing in Bangla", Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 642-647(2004).
- [3] K. D. Islam, M. Billah, R. Hasan and M. M. Asaduzzaman, "Syntactic Transfer and Generation of Complex-Compound Sentences for Bangla-English Machine Translation", Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 321-326 (2003).
- [4] L. Mehedy, N. Arifin and M. Kaykobad, "Bangla Syntax Analysis: A Comprehensive Approach", Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 287-293 (2003).
- [5] S. K. Chakravarty, K. Hasan, A. Alim, "A Machine Translation (MT) Approach to Translate Bangla Complex Sentences into English" Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 342-346 (2003).
- [6] S. A. Rahman, K. S. Mahmud, B. Roy and K. M. A. Hasan, "English to Bengali Translation Using A New Natural Language Processing Algorithm" Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 294-298 (2003).
- [7] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2<sup>nd</sup> Edition, Pearson Education publisher, New York, 2003.
- [8] M. M. Anwar, M. Z. Anwar and M. A. Bhuiyan, "Syntax Analysis and Machine Translation of Bangla Sentences", IJCSNS International Journal of Computer Science and

Network Security, VOL.9 No.8, August 2009, pp.  
317–326 (2009).



**Md. Musfique Anwar** completed his B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, Jahangirnagar University, Bangladesh in 2006. He is now a Lecturer in

the Dept. of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh. His research interests include Natural Language Processing, Artificial Intelligence, Image Processing, Pattern Recognition, Software Engineering and so on.



**Nasrin Sultana Shume** completed her B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, Jahangirnagar University, Bangladesh in 2006. She is now a Lecturer

in the Dept. of CSE, Green University of Bangladesh, Mirpur, Dhaka, Bangladesh. Her research interests include Artificial Intelligence, Neural Networks, Image Processing, Pattern Recognition, Database and so on.



**Md. Al-Amin Bhuiyan** received his B.Sc (Hons) and M.Sc. in Applied Physics and Electronics from University of Dhaka, Dhaka, Bangladesh in 1987 and 1988, respectively. He got the Dr. Eng. Degree in Electrical

Engineering from Osaka City University, Japan, in 2001. He has completed his Postdoctoral in the Intelligent Systems from National Informatics Institute, Japan. He is now a Professor in the Dept. of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh. His main research interests include Image Face Recognition, Cognitive Science, Image Processing, Computer Graphics, Pattern Recognition, Neural Networks, Human-machine Interface, Artificial Intelligence, Robotics and so on.

# Image Retrieval using Shape Texture Patterns generated from Walsh-Hadamard Transform and Gradient Image Bitmaps

Dr. H.B.Kekre<sup>1</sup>, Sudeep D. Thepade<sup>2</sup>, Varun K. Banura<sup>3</sup>

<sup>1</sup>Senior Professor, <sup>2</sup>Ph.D.Research Scholar & Associate Professor, <sup>3</sup>B.Tech (CE) Student  
Computer Engineering Department, MPSTME,

SVKM's NMIMS (Deemed-to-be University), Mumbai, India

<sup>1</sup>hbkekke@yahoo.com, <sup>2</sup>sudeepthepade@gmail.com, <sup>3</sup>varunkbanura@gmail.com

**Abstract**— The theme of the work presented here is gradient mask texture based image retrieval techniques using image bitmaps and texture patterns generated using Walsh-Hadamard transform. The shape of the image is extracted by using three different gradient operators (Prewitt, Robert and Sobel) with slope magnitude method followed by generation of bitmap of the shape feature extracted. This bitmap is then compared with the different texture patterns namely '4-pattern', '16-pattern' and '64-pattern' generated using Walsh-Hadamard transform matrix to produce the feature vector as the matching number of ones and minus ones per texture pattern. The proposed content based image retrieval (CBIR) techniques are tested on a generic image database having 1000 images spread across 11 categories. For each proposed CBIR technique 55 queries (randomly selected 5 per image category) are fired on the image database. To compare the performance of image retrieval techniques average precision and recall of all the queries per image retrieval technique are computed. The results have shown improved performance (higher precision and recall values of crossover points) with the proposed methods compared to the mask-shape based image retrieval techniques. Further the performance of proposed image retrieval methods is enhanced using even image part. In the discussed image retrieval methods, the combination of original and even image part for 4-pattern texture with shape masks generated using Robert gradient operator gives the highest crossover point of precision and recall indicating better performance.

**Keywords**- CBIR, Gradient operators, Walsh-Hadamard transform, Texture, Pattern, Bitmap.

## I. INTRODUCTION

Today the information technology experts are facing technical challenges to store/transmit and index/manage image data effectively to make easy access to the image collections of tremendous size being generated due to large numbers of images generated from a variety of sources (digital camera, digital video, scanner, the internet etc.). The storage and transmission is taken care of by image compression [4,7,8]. The image indexing is studied in the perspective of image database [5,9,10,13,14] as one of the promising and important research area for researchers from disciplines like computer vision, image processing and database areas. The hunger of superior and quicker image retrieval techniques is increasing day by day. The significant applications for CBIR technology could be listed as art galleries [15,17], museums, archaeology

[6], architecture design [11,16], geographic information systems [8], weather forecast [8,25], medical imaging [8,21], trademark databases [24,26], criminal investigations [27,28], image search on the Internet [12,22,23]. The paper attempts to provide better and faster image retrieval techniques.

### A. Content Based Image Retrieval

For the first time Kato et.al. [7] described the experiments of automatic retrieval of images from a database by colour and shape feature using the terminology content based image retrieval (CBIR). The typical CBIR system performs two major tasks [19,20] as feature extraction (FE), where a set of features called feature vector is generated to accurately represent the content of each image in the database and similarity measurement (SM), where a distance between the query image and each image in the database using their feature vectors is used to retrieve the top "closest" images [19,20,29].

For feature extraction in CBIR there are mainly two approaches [8] feature extraction in spatial domain and feature extraction in transform domain. The feature extraction in spatial domain includes the CBIR techniques based on histograms [8], BTC [4,5,19], VQ [24,28,29]. The transform domain methods are widely used in image compression, as they give high energy compaction in transformed image [20,27]. So it is obvious to use images in transformed domain for feature extraction in CBIR [26]. But taking transform of image is time consuming. Spatial feature based CBIR methods are given in [30] as mask-shape CBIR and mask-shape BTC CBIR. The proposed CBIR methods are further attempting to improve the performance of these shape based image retrieval with help of shape texture patterns. Here the query execution time is further reduced by decreasing the feature vector size further and making it independent of image size. Many current CBIR systems use the Euclidean distance [4-6,11-17] on the extracted feature set as a similarity measure. The Direct Euclidean Distance between image P and query image Q can be given as equation 1, where V<sub>pi</sub> and V<sub>qi</sub> are the feature vectors of image P and Query image Q respectively with size 'n'.

$$ED = \sqrt{\sum_{i=1}^n (V_{pi} - V_{qi})^2} \quad (1)$$



## II. EDGE DETECTION MASKS

Edge detection is a very important in image analysis. As the edges give idea about the shapes of objects present in the image so they are useful for segmentation, registration, and identification of objects in a scene. An edge is a jump in intensity. An ideal edge is a discontinuity (i.e., a ramp with an infinite slope). The first derivative assumes a local maximum at an edge. The various gradient operators [13] used for edge extraction are Prewitt, Roberts and Sobel.

## III. SLOPE MAGNITUDE METHOD

The problem with edge extraction using gradient operators is detection of edges in only either horizontal or vertical directions. Shape feature extraction in image retrieval requires the extracted edges to be connected in order to reflect the boundaries of objects present in the image. Slope magnitude method is used along with the gradient operators (Prewitt, Robert and Sobel) to extract the shape features in form of connected boundaries. The process of applying the slope magnitude method is given as follows. First one needs to convolve the original image with the Gx mask to get the x gradient and Gy mask to get the y gradient of the image. Then the individual squares of both are taken. Finally the two squared terms are added and square root of this sum is taken as given in equation 2.

$$G = \sqrt{G_x^2 + G_y^2} \quad (2)$$

## IV. TEXTURE PATTERNS USING WALSH-HADAMARD TRANSFORM MATRIX

Walsh transform matrix [21,22,26] is defined as a set of N rows, denoted W<sub>j</sub>, for j = 0, 1, ..., N - 1, which have the following properties:

- W<sub>j</sub> takes on the values +1 and -1.
- W<sub>j</sub>[0] = 1 for all j.
- W<sub>j</sub>W<sub>k</sub><sup>T</sup>=0, for j not equal to k and W<sub>j</sub>W<sub>k</sub><sup>T</sup>=N, for j=k.
- W<sub>j</sub> has exactly j zero crossings, for j = 0, 1, ..., N-1.
- Each row W<sub>j</sub> is even or odd with respect to its midpoint

Walsh transform matrix is defined using a Hadamard matrix of order N. The Walsh transform matrix row is the row of the Hadamard matrix specified by the Walsh code index, which must be an integer in the range [0, ..., N - 1]. For the Walsh code index equal to an integer j, the respective Hadamard output code has exactly j zero crossings, for j = 0, 1, ..., N - 1.

Using the Walsh-Hadamard transform assorted texture patterns namely 4-pattern, 16-pattern and 64-pattern are generated. To generate N<sup>2</sup> texture patterns, each column of the Walsh-Hadamard matrix of size NxN is multiplied with every element of all possible columns of the same matrix (one

column at a time to get one pattern). The texture patterns obtained are orthogonal in nature.

Figure 1(a) shows a 2X2 Walsh-Hadamard matrix. The four texture patterns generated using this matrix are shown in figure 1(b). Similarly figure 2(b) shows first four texture patterns (out of total 16) generated using 4X4 Walsh-Hadamard matrix shown in figure 2(a).

1	1
1	-1

1(a). 2x2 Walsh-Hadamard transform matrix

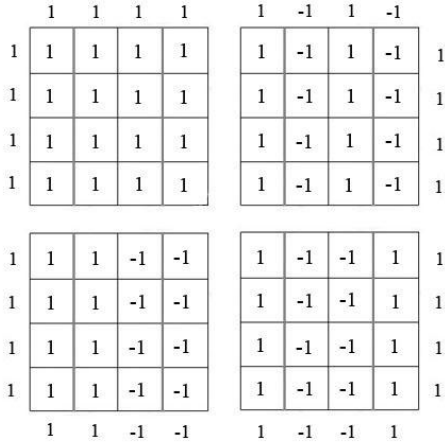
	1	1	1	1	
1	1	1	1	1	1
1	1	1	-1	-1	-1
1	1	-1	1	-1	1
1	1	-1	-1	1	-1
	1	-1	1	-1	

1(b). '4-pattern' texture patterns

Figure 1. Generating 4 texture patterns

1	1	1	1
1	-1	1	-1
1	1	-1	-1
1	-1	-1	1

2(a). 4x4 Walsh-Hadamard transform matrix



2(b). First four '16-pattern' texture patterns

Figure 2. Generating first four of 16 texture patterns

#### V. GRADIENT IMAGE BITMAPS

Image bitmaps of colour image are generated using three independent red (R), green (G) and blue (B) components of Prewitt/Robert/Sobel image obtained using slope magnitude method to calculate three different thresholds. Let  $X=\{R(i,j),G(i,j),B(i,j)\}$  where  $i=1,2,\dots,m$  and  $j=1,2,\dots,n$ ; be an  $m \times n$  slope magnitude gradient of color image in RGB space. Let the thresholds be TR, TG and TB, which could be computed as per the equations given below as 3, 4 & 5.

$$TR = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n R(i, j) \quad (3)$$

$$TG = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n G(i, j) \quad (4)$$

$$TB = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n B(i, j) \quad (5)$$

Here three binary bitmaps will be computed as BMr, BMg and BMb. If a pixel in each component (R, G, and B) is greater than or equal to the respective threshold, the corresponding pixel position of the bitmap will have a value of 1 otherwise it will have a value of -1.

$$BMr(i, j) = \begin{cases} 1, & \text{if } \dots R(i, j) \geq TR \\ -1, & \dots \text{if } \dots R(i, j) < TR \end{cases} \quad (6)$$

$$BMg(i, j) = \begin{cases} 1, & \text{if } \dots G(i, j) \geq TG \\ -1, & \dots \text{if } \dots G(i, j) < TG \end{cases} \quad (7)$$

$$BMb(i, j) = \begin{cases} 1, & \text{if } \dots B(i, j) \geq TB \\ -1, & \dots \text{if } \dots B(i, j) < TB \end{cases} \quad (8)$$

To generate tiled bitmaps, the image is divided into four non-overlapping equal quadrants and the average of each quadrant is considered to generate the respective tile of the image bitmap.

#### VI. DISCUSSED CBIR METHODS

##### A. Mask-shape based CBIR

In this method the feature vector is obtained by extracting the shape of the image by using gradient operators (Prewitt, Robert or Sobel) with slope magnitude method. Then the feature vectors are compared pixel by pixel using the Euclidian distance. The limitation of this method is that it is dependent on the size of the image. For this method, size of all the images in the database should be same as query image.

##### B. CBIR with Mask-shape and BTC

First of all the shape of the image is extracted by using three gradient operators with slope magnitude method. The average of the obtained shape feature is calculated. The feature vector is obtained by calculating the average of all those values which are greater than the average of the shape feature and average of all those values which are less than or equal to the average of the shape feature. Here the size of the feature vector is constant and is independent of size of the image.

##### C. Proposed Shape Texture Pattern based CBIR

In the proposed gradient shape texture method, the shape feature of the image is extracted using the three gradient operators Prewitt, Robert and Sobel. Then the bitmap of the shape feature is generated using the modified BTC technique. The bitmap thus obtained is compared with the different texture patterns like '4-pattern', '16-pattern' and '64-pattern' generated using Walsh-Hadamard transform matrix to produce the feature vector as the matching number of ones and minus ones per texture pattern. The size of the feature vector of the image is given by equation 8.

$$\text{Feature vector size} = 2 * 3 * (\text{no. of considered texture-pattern}) \quad (9)$$

Using three different gradient operators in association with three assorted texture pattern sets along with original and original-even image, total 18 novel feature vector generation methods can be used resulting into 18 new image retrieval techniques. Mask-shape based CBIR methods [30] are considered to compare the performance of proposed CBIR techniques. In the proposed CBIR techniques the combination of original and even part of images gives better results than original image alone [1,2]. The main advantage of proposed CBIR methods is improved performance resulting into better image retrieval. Here also the feature vector size is independent of image size in proposed CBIR methods.

Table 1. Feature vector size of discussed image retrieval techniques

CBIR Technique	Feature vector size for NxN image	CBIR Technique	Feature vector size for NxN image
Mask-shape	NxN	4-Pattern	8
Mask-Shape + BTC	2	16-Pattern	32
		64-Pattern	128

## VII. IMPLEMENTATION

The implementation of the discussed CBIR techniques is done in MATLAB 7.0 using a computer with Intel Core 2 Duo Processor T8100 (2.1GHz) and 2 GB RAM.

The CBIR techniques are tested on the Wang image database [18] of 1000 variable size images spread across 11 categories of human being, animals, natural scenery and manmade things, etc. The categories and distribution of the images is shown in table 2.

Table 1. Image Database: Category-wise Distribution

Category	Tribes	Buses	Beaches
No. of Images	85	99	99
Category	Horses	Mountains	Airplanes
No. of Images	99	61	100
Category	Dinosaurs	Elephants	Roses
No. of Images	99	99	99
Category	Monuments	Sunrise	
No. of Images	99	61	

To assess the retrieval effectiveness, we have used the precision and recall as statistical comparison parameters [4,5] for the proposed CBIR techniques. The standard definitions for these two measures are given by the equations 10 and 11.

$$\text{Precision} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_images\_retrieved}} \quad (10)$$

$$\text{Recall} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_relevant\_images\_in\_database}} \quad (11)$$

## VIII. RESULTS AND DISCUSSION

For testing the performance of each proposed CBIR method, 55 queries (randomly selected 5 from each image

category) are fired on the image database. The feature vector of query image and database image are matched using the Euclidian distance. The average precision and recall values are found for all the proposed CBIR methods. The intersection of precision and recall values gives the crossover point. The crossover point of precision and recall is computed for all the proposed CBIR methods. The one with higher value of crossover point indicates better performance.

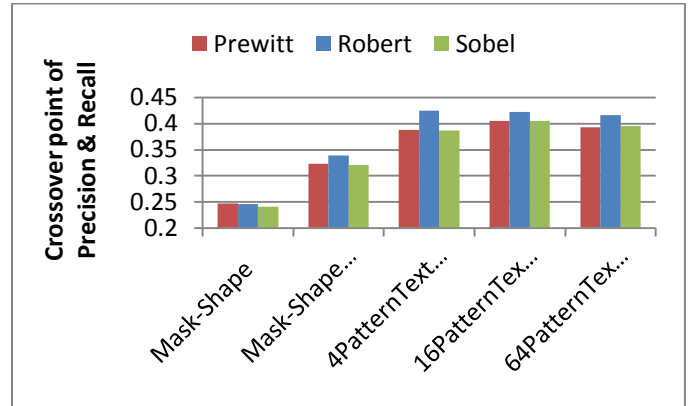


Figure 3. Performance comparison of proposed CBIR methods with the mask-shape based CBIR methods

Figure 3 shows the performance comparison of proposed CBIR methods with the mask-shape based CBIR methods [30]. It is observed that the 'mask-shape' based image retrieval gives the worst performance. However the '4-pattern' texture based image retrieval with Robert as gradient operator has the highest crossover point thus indicating better performance. It is also observed that in case of 'mask-shape' based image retrieval technique, Prewitt operator gives the best results followed by Robert and Sobel. On the other hand Robert outperforms the other two operators in case of mask-shape texture based image retrieval techniques.

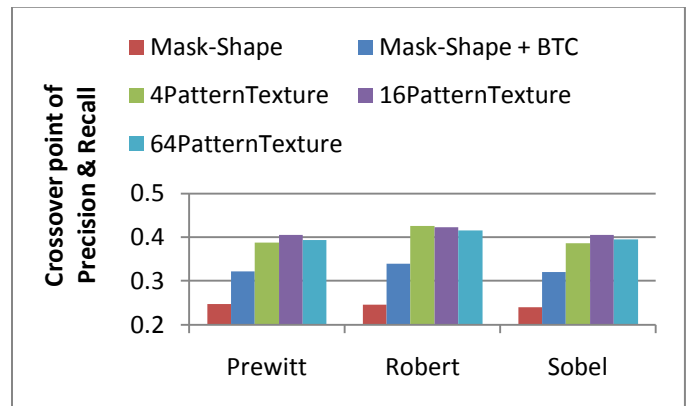


Figure 4. Performance comparison of the proposed CBIR methods with different gradient operators

Figure 4 shows performance comparison of the proposed CBIR methods with different gradient operators namely Prewitt, Robert and Sobel. In case of Prewitt and Sobel, as the number of texture patterns are increased (upto '16-pattern' texture) the value of precision-recall crossover point also increases. Beyond '16-pattern' texture, the results start

deteriorating. In case of Robert, '4-Pattern' texture gives the highest crossover point and on increasing the number of texture patterns, the results start degrading.

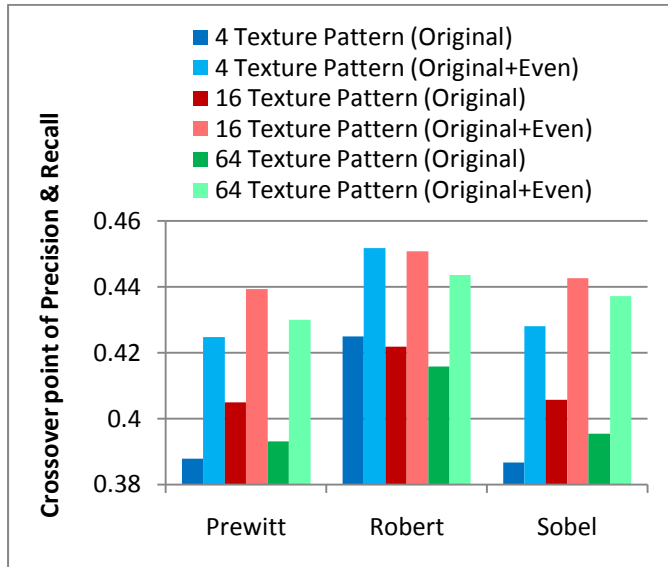


Figure 5. Amelioration of the proposed CBIR method with the combination of original and even image part using different gradient operators

Figure 5 shows amelioration of the proposed CBIR methods with the combination of original and even image part using different gradient operators namely Prewitt, Robert and Sobel. It is observed that in case of Prewitt and Sobel operators, the combination of original with even image part for '16-pattern' texture gives the highest precision-recall crossover point. In case of Robert operator, the combination of original with even image part for '4-pattern' texture gives the best performance.

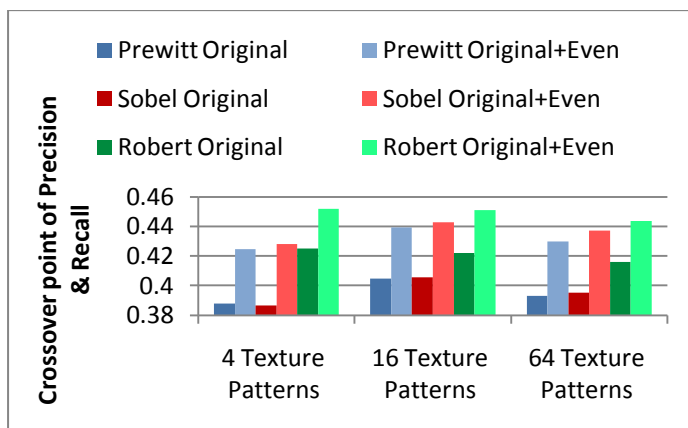


Figure 6. Performance comparison of the proposed CBIR methods using the combination of original and even image part

Figure 6 shows the performance comparison of the proposed CBIR methods using the combination of original and even image part. It is observed that the combination of original with even image part gives better performance than original alone.

In all the three types of texture based image retrieval using different gradient operators, Robert outperforms the other two gradient operators Prewitt and Sobel.

From comparison of the mask-shape based image retrieval techniques [30] with the proposed CBIR methods it is observed that the '4-pattern' texture based image retrieval with the combination of original with even image part using Robert gradient operator gives the best result. In case of Prewitt and Sobel operators, increasing the number of texture patterns ameliorates the image retrieval performance upto a certain level ('16-pattern' texture) beyond which the results start deteriorating.

## IX. CONCLUSION

Better image retrieval techniques are proposed using shape texture patterns generated with help of Walsh-Hadamard transform and gradient image bitmaps. Among the gradient operators "Robert" proved to be better for CBIR in proposed methods. For Robert, '4-pattern' has given best performance and '16-pattern' is proven to be better in Sobel and Prewitt operators. In all Robert 4-pattern has given best performance. The methods have been ameliorated using even image part along with original one for further improvement in CBIR performance.

## X. REFERENCES

- [1] Dr. H.B.Kekre, Sudeep D. Thepade, Varun K. Banura, "Augmentation of Colour Averaging Based Image Retrieval Techniques using Even part of Images and Amalgamation of feature vectors", International Journal of Engineering Science and Technology (IJEST), Volume 2, Issue 10, (ISSN: 0975-5462) Available online at <http://www.ijest.info>
- [2] Dr. H.B.Kekre, Sudeep D. Thepade, Varun K. Banura, "Amelioration of Colour Averaging Based Image Retrieval Techniques using Even and Odd parts of Images", International Journal of Engineering Science and Technology (IJEST), Volume 2, Issue 9, (ISSN: 0975-5462) Available online at <http://www.ijest.info>.
- [3] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo, "Query by Image Content Using Colour Averaging Techniques", International Journal of Engineering Science and Technology (IJEST), Volume 2, Issue 6, 2010, pp.1612-1622 (ISSN: 0975-5462) Available online at <http://www.ijest.info>.
- [4] Dr. H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECE), Volume 2, Number 3, pp. 172-180, Summer 2008. Available online at <http://www.waset.org/ijecse/v2/v2-3-23.pdf>
- [5] Dr. H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigues College of Engg., Mumbai. Is uploaded on online ACM portal.
- [6] Dr. H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html), Volume 1, No. A08, pp. 31-46, Autumn 2008.
- [7] Hirata K. and Kato T. "Query by visual example – content-based image retrieval", In Proc. of Third International Conference on Extending Database Technology, EDBT'92, 1992, pp 56-71
- [8] Dr. H.B.Kekre, Sudeep D. Thepade, "Rendering Futuristic Image Retrieval System", National Conference on Enhancements in Computer, Communication and Information Technology, EC2IT-2009, 20-21 Mar

- 2009, K.J.Somaiya College of Engineering, Vidyavihar, Mumbai-77.
- [9] Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Transactions On Image Processing, Volume 11, Number 2, pp.146-158, February 2002.
- [10] B.G.Prasad, K.K. Biswas, and S. K. Gupta, "Region -based image retrieval using integrated color, shape, and location index", International Journal on Computer Vision and Image Understanding Special Issue: Colour for Image Indexing and Retrieval, Volume 94, Issues 1-3, April-June 2004, pp.193-233.
- [11] Dr. H.B.Kekre, Sudeep D. Thepade, "Creating the Color Panoramic View using Medley of Grayscale and Color Partial Images ", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No. 3, Summer 2008. Available online at [www.waset.org/ijecse/v2/v2-3-26.pdf](http://www.waset.org/ijecse/v2/v2-3-26.pdf).
- [12] Stian Edvardsen, "Classification of Images using color, CBIR Distance Measures and Genetic Programming", Ph.D. Thesis, Master of science in Informatics, Norwegian university of science and Technology, Department of computer and Information science, June 2006.
- [13] R.C. Gonsalez, R.E. Woods "Digital Image Processing", Second Edition, Pearson Publication
- [14] Zhibin Pan, Kotani K., Ohmi T., "Enhanced fast encoding method for vector quantization by finding an optimally-ordered Walsh transform kernel", ICIP 2005, IEEE International Conference, Volume 1, pp 1 - 573-6, Sept. 2005.
- [15] Dr. H.B.Kekre, Sudeep D. Thepade, "Improving 'Color to Gray and Back' using Kekre's LUV Color Space", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009. Is uploaded at online at IEEE Xplore.
- [16] Dr. H.B.Kekre, Sudeep D. Thepade, "Image Blending in Vista Creation using Kekre's LUV Color Space", SPIT-IEEE Colloquium and International Conference, Sardar Patel Institute of Technology, Andheri, Mumbai, 04-05 Feb 2008.
- [17] Dr. H.B.Kekre, Sudeep D. Thepade, "Color Traits Transfer to Grayscale Images", In Proc.of IEEE First International Conference on Emerging Trends in Engg. & Technology, (ICETET-08), G.H.Raisoni COE, Nagpur, INDIA. Uploaded on online IEEE Xplore.
- [18] <http://wang.ist.psu.edu/docs/related/Image.orig> (Last referred on 23 Sept 2008)
- [19] Dr. H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.
- [20] Dr. H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS), Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at [www.IJCSNS.org](http://www.IJCSNS.org).
- [21] Dr. H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Walsh Transform over Row Mean and Column Mean using Image Fragmentation and Energy Compaction for Image Retrieval", International Journal on Computer Science and Engineering (IJCSE), Volume 2S, Issue1, January 2010, (ISSN: 0975-3397). Available online at [www.eggjournals.com/ijcse](http://www.eggjournals.com/ijcse).
- [22] Dr. H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue I, Feb.2010, pp.9-18, Available online [www.icgst.com/gvip/Volume10/Issue1/P1150938876.html](http://www.icgst.com/gvip/Volume10/Issue1/P1150938876.html)
- [23] Dr. H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 2-14. Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html).
- [24] Dr. H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 55-65. Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html) (ISSN: 0974-0627).
- [25] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Performance Comparison for Face Recognition using PCA, DCT & Walsh Transform of Row Mean and Column Mean", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue II, Jun.2010, pp.9-18, Available online <http://209.61.248.177/gvip/Volume10/Issue2/P1181012028.pdf>.
- [26] Dr. H.B.Kekre, Sudeep D. Thepade, "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image", International Journal of Information Retrieval, Serials Publications, Volume 2, Issue 1, 2009, pp. 72-79 (ISSN: 0974-6285)
- [27] Dr. H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.
- [28] Dr. H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi, "Improved Texture Feature Based Image Retrieval using Kekre's Fast Codebook Generation Algorithm", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.
- [29] Dr. H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval by Kekre's Transform Applied on Each Row of Walsh Transformed VQ Codebook", (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is invited at ICWET 2010. Also will be uploaded on online ACM Portal.
- [30] Dr. H.B.Kekre, Sudeep D. Thepade, Shobhit W., P. Mukharjee, S. Singh, Miti K., "Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC", International Journal of Computer Applications (IJCA), Volume 6, Number 8, pp.28-23, September 2010. Available online at <http://www.ijcaonline.org/volume6/number8/pxc3871430.pdf>

#### AUTHORS PROFILE



**Dr. H. B. Kekre** has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering. College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS University. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./B.Tech projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 320 papers in National / International Conferences and Journals to his credit. He was Senior Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE Recently ten students working under his guidance have received best paper awards and two have been conferred Ph.D. degree of SVKM's NMIMS University. Currently 10 research scholars are pursuing Ph.D. program under his guidance.





**Sudeep D. Thepade** has Received B.E.(Computer) degree from North Maharashtra University with Distinction in 2003. M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, currently pursuing Ph.D. from SVKM's NMIMS, Mumbai. He has about than 07 years of experience in teaching and industry. He was Lecturer in Dept. of Information Technology at Thadomal Shahani Engineering College, Bandra(w), Mumbai for nearly 04 years. Currently working as Associate Professor in Computer Engineering at Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. He has been on International Advisory Board of many International Conferences. He is Reviewer for many reputed International Journals. His areas of interest are Image Processing and Computer Networks. He has about 100 papers in National/International Conferences/Journals to his credit with a Best Paper Award at International Conference SSPCCIN-2008, Second Best Paper Award at ThinkQuest-2009 National Level paper presentation competition for faculty and Best Paper Award at Springer International Conference ICCCT-2010.



**Varun K. Banura** is currently pursuing B.Tech. (CE) from MPSTME, NMIMS University, Mumbai. His areas of interest are Image Processing and Computer Networks. He has 05 research papers in International Conferences/Journals to his credit.

# An Evaluation of Software Requirement Prioritization Techniques

Mohammad Shabbir Hasan<sup>1, a</sup>, Abdullah Al Mahmood<sup>2, a</sup>, Md. Jahangir Alam<sup>3, b</sup>, Sk. Md. Nahid Hasan<sup>4, a</sup>, Farin Rahman<sup>5, a</sup>

E-mail: <sup>1</sup>shabbir\_cse03@yahoo.com, <sup>2</sup>sajibcseofkuet@gmail.com, <sup>3</sup>palash\_14@yahoo.com,  
<sup>4</sup>auvi\_2k5@hotmail.com, <sup>5</sup>farinrahman.cse@gmail.com,

a. Panacea Research Lab, Dhaka, Bangladesh

b. Dhaka International University, Dhaka, Bangladesh

**Abstract**— Requirements prioritization plays an important role in the requirement engineering process, particularly, with respect to critical tasks like requirements negotiation and software release planning. Selecting the right set of requirements for a product release largely depends on how successfully the requirement candidates are prioritized. There are different requirement prioritization techniques available which are some more elaborated than others. This paper takes a closer look at nine different techniques of requirement prioritization namely Analytical Hierarchy Process (AHP), Hierarchy AHP, Minimal Spanning Tree, Bubble Sort, Binary Search Tree (BST), Priority Group, Planning Game (PG), 100 points method and Planning Game combined with AHP (PGcAHP) and then put them into a controlled experiment, in order to find out the best one. The evaluation was done on the basis of some criteria like: ease of use, certainty, accuracy of result, method's ability to scale up to many more requirements, required number of comparisons, and required time to make decision. Analysis of the data from the experiment indicates that the analytic hierarchy process to be a promising candidate, although it may be problematic to scale-up. However, the result clearly indicates that the Planning Game (PG) yields accurate result, is able to scale up, requires least amount of time, the easiest method to use and so on. For these reasons, finding of the experiment is, the Planning Game (PG) method is supposed to be the best method for prioritizing requirements.

**Keywords**- Requirement Engineering, Requirement Prioritization, Requirement Negotiation, Software Product Management, Software Release Planning.

## I. INTRODUCTION

In market-driven software development, products which are intended for an open market are developed in several consecutive releases. Market-driven development does not have easily identifiable customers and the requirements often need to be invented based on the needs of several potential users [1]. When only one stakeholder is involved in the project, it is relatively easy to make decisions since only one stakeholder's opinion needs to be considered. But when more than one stakeholder is involved in the project, decisions can be harder to make, since different stakeholders

have different perspectives and in this case, more requirements are yield than can be implemented at once. Again, not all the requirements contain equal user satisfaction. For example, project developers look for the requirements which can be implemented fast, financial managers look for the requirements with low cost, market managers look for the requirements with high market value, and end users look for the requirements which are easy to use. One requirement may be of low cost, with short implementation time, but also have low market value and be hard to use. Conversely, another requirement may have a high cost, but short time to be implemented, high market value and be easy to use. It can be a challenge for the software development team to decide which requirements need to be implemented first. Requirements prioritization is a technique that can uncover the most important requirements to maximize the stakeholders' satisfaction.

In commercial software development, decision makers need to make many different decisions regarding the release plan according to some issues like available resources, milestones, conflicting stakeholder views. Available market opportunity, risks, product strategies, and costs need to be taken into consideration when planning future releases. Nowadays, unfortunately, projects are suffering low success rates. According to an annual report named 'CHAOS Summary 2009' prepared by Standish Group [27], only 32% of all projects were considered as successful which are delivered on time, on budget, with required features and functions. Among the rest, 44% were challenged which are late, over budget, and/or with less than the required features and functions and 24% failed which are cancelled prior to completion or delivered and never used. Ten main factors causing challenged or failed projects are unveiled. Four of them are lack of user involvement, lack of resources, unrealistic expectations, and changing requirements and specifications. Requirements prioritization increases user involvement by letting the stakeholders decide which requirements the project should contain. It helps stakeholders to be realistic by letting them understand the current constraints on resources and accepting the trade-off

decisions on conflicting perspectives. Karlsson et al think it helps stakeholders to allocate resources based on the priorities of the requirements [2], detect requirements defects, such as misunderstanding or ambiguous requirements [3] and reduce the number of changes to requirements and specifications in the later stage of projects. Hatton [4] says requirements prioritization has become an essential step in the software development process in order to reduce software failure. Ngo-The and Ruhe [5] note requirements prioritization has been recognized as one of the most important decision making processes in the software development process.

Several approaches have been proposed [6-10] which adopts a common model for the requirements prioritization process. This paper provides an investigation of nine candidate methods for prioritizing requirements: Analytic Hierarchy Process (AHP), Hierarchy AHP, Minimal Spanning Tree, Bubble Sort, Binary Search Tree (BST), Planning Game (PG), Priority Group, 100 points method and Planning Game Combined with AHP (PGcAHP). To study these methods, we systematically applied all methods to prioritize 14 well-defined quality requirements of a mobile set. We then categorized the methods from a user's perspective according to a number of criteria such as accuracy, certainty, method's ability to scale up to many more requirements, required time to make decision, total number of decisions and ease of use.

This paper is organized as follows. Section 2 motivates this work, and the paper continues in Section 3 by outlining the nine different prioritizing methods. Section 4 describes the evaluation framework and Section 5 presents the way to find out the best one among the techniques under consideration followed by a discussion of the result in Section 6. We finish by drawing some broad and necessarily speculative and personal conclusions in Section 7.

## II. MOTIVATION

Industrial software development has a growing acknowledgement that requirements are of varying importance. Yet there has been little progress to date, either theoretical or practical, on the mechanisms for prioritizing software requirements [11]. In a review of the state of the practice in requirements engineering, Lubars et al. [12] found that many organizations believe that it is important to assign priorities to requirements and to make decisions about them according to rational, quantitative data. Still it appeared that no company really knew how to assign priorities or how to communicate these priorities effectively to project members [3].

A sound basis for prioritizing software requirements is the approach provided by the analytic hierarchy process, AHP [13] where decision makers compare the requirements pair-wise to determine which of the two is more important, and to what extent. In industrial projects, this approach has been experienced as being effective, accurate and also to yield informative and trustworthy results [7]. Probably even

more important, after using the approach in several commercial projects, practitioners are found to be very attracted by the approach, and continue to use it in other projects [3]. AHP has only been used in few applications in the software industry. Finnie et al. [14], for example, used AHP to prioritize software development factors. Other applications of AHP include a telecommunications quality study performed by Douligeris and Pereira [15], and software requirements prioritizing in a commercial development project by Karlsson [16]. Despite some positive experience, AHP has a fundamental drawback which impedes its industrial institutionalization. Since all unique pairs of requirements are to be compared, the required effort can be substantial. In small-scale development projects this growth rate may be acceptable, but in large-scale development projects the required effort is most likely to be overwhelming [3].

Since AHP may be problematic for large-scale projects, Karlsson et al [3] identified five complementary approaches to challenge AHP. All of these methods involve pair-wise comparisons, since previous studies indicate that making relative judgments tends to be faster and still yield more reliable results than making absolute judgments [7]. They focused on methods which may reduce the required effort, but still able to produce high-quality results, considered trustworthy by its users. Again, Paetsch et al [17] claims that agile software development has become popular during the last few years and in this field, one of the most popular methods is the extreme programming, which has a prioritization technique called Planning Game (PG). In this paper, we investigated PG with all the requirement prioritization techniques used in the experiment carried out by Karlsson et al [3]. We also investigated a rather easy and quick method (at least according to the theory), and that is the 100 points method. Next section gives a brief description of each method, both in theory and then how it works practically.

## III. PRIORITIZATION METHODS

Prioritizing methods guide decision makers to analyze requirements to assign numbers or symbols that reflect their importance. According to Karlsson et al [3], a prioritizing session may consist of three consecutive stages:

- (1) The *preparation* stage where a person structures the requirements according to the principle of the prioritizing methods to be used. A team and a team leader for the session is selected and provided all necessary information.
- (2) The *execution* stage where the decision makers do the actual prioritizing of the requirements using the information they were provided with in the previous stage. The evaluation criteria must be agreed upon by the team before the execution stage is initiated.

- (3) The *presentation* stage where the results of the execution are presented for those involved. Some prioritizing methods involve different kinds of calculations that must be carried out before the results can be presented.

This section describes the prioritization techniques investigated in this paper:

#### A. Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) was first developed and explained by Saaty [13] in 1980. Regnell et al [18] claim that even though this is a promising technique, the technique itself is not adapted to distributed prioritization with multiple stakeholders; hence it has to be modified in one way or another. However, at present time there have not been published any research how that kind of modification would function.

In AHP the candidate requirements are compared pairwise, and to which extent one of the requirements is more important than the other requirement. Saaty [13] states that the intensity of importance should be according to Table 1.

TABLE I. BASIC SCALE ACCORDING TO SAATY [13] FOR PAIR-WISE COMPARISONS IN AHP

How Important	Description
1	Equal importance
3	Moderate difference in importance
5	Essential difference in importance
7	Major difference in importance
9	Extreme difference in importance
Reciprocals	If requirement <i>i</i> has one of the above numbers assigned to it when compared with requirement <i>j</i> , then <i>j</i> has the reciprocal value when compared with <i>i</i> .

To fully understand AHP it is easiest to divide AHP into three different phases.

1. Comparing every pair of requirements, this is the “engine” of AHP, according to Saaty [19].
2. Derives a priority vector of relative weights for these requirements, i.e. the principal eigenvector.
3. Calculate the by-product from 2, i.e. the inconsistency measure.

First we take the requirements that should be prioritized (the total amount of requirement is *n*), and put them into a matrix, where the rows have the index of *i* and columns have the index of *j*. The matrix is called *W* and the elements in the matrix are called *w*. The requirement that is placed in row *i* and column *j* gets the index *ij*. Therefore the element *w<sub>ij</sub>* has the row index = *i* and column index = *j*.

TABLE II. MATRIX OF PAIR WISE COMPARISONS

	Requirement 1	Requirement 2	.....	Requirement n
Requirement 1	1	<i>w<sub>12</sub></i>	<i>w<sub>1j</sub></i>	<i>w<sub>1n</sub></i>
Requirement	<i>w<sub>21</sub></i>	1	<i>w<sub>2j</sub></i>	<i>w<sub>2n</sub></i>

2				
.....	<i>w<sub>j1</sub></i>	<i>w<sub>j2</sub></i>	1	<i>w<sub>jn</sub></i>
Requirement n	<i>w<sub>n1</sub></i>	<i>w<sub>n2</sub></i>	<i>w<sub>nj</sub></i>	1

Each matrix element consists of the comparison between two requirements (*i* and *j*), which gives us the following relationship:

$$w_{ij} = \frac{w_i}{w_j} \quad (1)$$

An important notice is that the person that does the prioritization does not put any value on *w<sub>i</sub>* and *w<sub>j</sub>*, instead he or she decides the value for *w<sub>ij</sub>* which is the ratio between *w<sub>i</sub>* and *w<sub>j</sub>*. That leads us to another important relationship, which is that for every index of *i, j, k* has the following relationship:

$$w_{ij} = \frac{w_i}{w_j}, w_{ij} = w_{ik} w_{kj} \quad (2)$$

With the information from formulae (1) and (2) and the matrix Table 2 we can see that some pair-wise comparisons are doing twice. The problem with human perception and judgments are subject to change if the human becomes tired or something changes the human psychological state (i.e. the level of blood sugar is dropping, and thereby the concentration). To solve this problem, Saaty [13] proposed that we should only compare *a<sub>ij</sub>*, *j* > *i*. With this solution we do not need to do *n*<sup>2</sup> comparison. Instead we only need to do half the comparison, since the formulae (2) say that *w<sub>ji</sub>* =  $\frac{1}{w_{ij}}$ . So it is really easy to apply this formula (2) to the comparisons that are not necessary. This leaves us to the diagonal, with the comparison with requirement *w<sub>i</sub>* and *w<sub>i</sub>* they will always be equal (i.e. the reciprocal value 1). Hence, we do not need to do this comparison either. This led us to the formulae (3):

$$\text{Total number of comparisons} = \frac{n(n-1)}{2} \quad (3)$$

The next step according to Saaty [13] is to calculate the eigenvector *v*. The elements of the eigenvector correspond to the priorities of the requirements. Gass and Rapcsák [20] describe it in the following way: If *W* is a consistent matrix, i.e. formulae (2) holds for all the indices of *i, j, k*, then *w* is of rank one and  $\lambda_{\max} = n$ . If the relationship  $\lambda_{\max} = n$  is true, *W* is a positive reciprocal matrix.

$$Wv = \lambda v \quad (4)$$

The formula (4) is the mathematical definition on the relationship between the eigenvalue and the eigenvector. This is nothing that is specific for AHP, but is valid for all matrices.

This means that  $v$  must be the eigenvector of  $W$  that correspond to the maximum eigenvalue  $\lambda$ . What this mean in reality is that every requirement is in the matrix and then the sum of  $j$  columns is calculated.

$$w_{11} + w_{21} + w_{31} + \dots + w_{(n-1)1} + w_{n1} = Z \quad (5)$$

Then each element in the column is divided by the sum,  $z$ , calculated from formulae (5). In the next step, elements in row  $i$  are added and then each row sum is divided by the total number of requirements  $n$ . This results a matrix representing the normalized eigenvector of the comparison matrix. Based on the elements in the normalized eigenvector, also known as the priority vector, conclusion about the importance of the requirements can be drawn. The degree of importance could be what kinds of business value that requirement yields or what it costs to develop or any other kind of importance. All depending on what aspect that the person that prioritized had in his/her mind during the prioritization.

The final step is to calculate how consistent the prioritization has been done. The reason of calculating this consistency is: If a person prioritizes that A is more important than B, B is more important than C and finally, C is more important than A, this will mean that C is more important than C, which cannot be true, i.e. the person has done a judgment error, hence it is important to find out if the person is consistent in his/her judgment. The consistency index (CI) is calculated by the formulae (6)

$$CI = \frac{\lambda_{\max} - n}{(n-1)} \quad (6)$$

$\lambda_{\max}$  is the maximum eigenvalue of the matrix. If the  $\lambda_{\max}$  value is close to  $n$ , the number of requirements, then there have been little judgment errors and the result is more consistent.

To check whether CI is acceptable, Consistency Ratio (CR) is calculated. The CR is a ratio from CI and RI, where RI is one of the Random Indices [13]. The value of RI can be obtained from Table 3:

TABLE III. RANDOM INDICES FOR AHP

1	2	3	4	5	6	7	8	9	10	11	12	13
0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.48	1.49	1.51	1.56

$$CR = \frac{CI}{RI}$$

(7)

According to Saaty [13] a result of CR within the range of 0.10 or less is to be considered acceptable.

All this calculation is the strong part of AHP, which give us a good way to evaluate whether the decision take is good or not. AHP has been used to predict a variety of decision, from stock market portfolio, economic forecasts, oil prices, political candidate, etc [3].

However AHP has some weakness. Since it takes  $\frac{n(n-1)}{2}$  comparisons, it does not work well for a large number of requirements. For example if the amount of requirement is 10, then it takes 45 comparisons. If the number of requirements is 20, then it takes 190 comparisons, and if the number of requirements is 100, it takes 4950 comparisons. A software project could have up to several thousand requirements, which mean that the number of comparisons exceeds over 500 000, and it will take too long time to compare all the requirements, and the question would then be; should AHP be used or not?

#### A-1. How to implement AHP in Requirement Prioritization:

- (1) As preparation, all unique pairs of requirements are outlined.
- (2) As execution, all outlined pairs of requirements compared using the scale in Table 1.
- (3) As presentation, the above mentioned method is used to estimate the relative priority of each requirement. The consistency ratio of the pair-wise comparisons is calculated using the methods mentioned above. The consistency ratio is an indicator of the reliability of the resulting priorities, and thus also an estimate of the judgmental errors in the pair-wise comparisons.

#### B. Hierarchy AHP

In large-scale development projects the requirements are often structured in a hierarchy of interrelated requirements [21]. The most generalized requirements are placed at the top of the hierarchy and the more specific requirements on levels below. Hierarchies are a common structure in daily use of AHP. But, to separate this hierarchical requirements structure from the flat requirements structure outlined previously, Karlsson et al [3] use the name *Hierarchy AHP*, we also use the same name in this paper.

#### B-1. How to implement Hierarchy AHP in Requirement Prioritization:

- (1) As preparation, all unique pairs of same level requirements are outlined in the hierarchy. Not all requirements are pair-wise compared to each other, but only those at the same level.



- (2) As execution, all outlined pairs of requirements are compared using the scale in Table 1.
- (3) As presentation, methods used for AHP are also applied at each level of the hierarchy. The priorities are then propagated down the hierarchy.

Hierarchy AHP possesses similar characteristics to AHP. Using a hierarchical structure reduces the required number of decisions, but also the amount of redundancy. Thus it is more sensitive to judgmental errors than AHP [3].

### C. Minimal Spanning Tree

The pair-wise comparisons in AHP provide interesting relationships to each other. For example, if requirement A is determined to be of higher priority than requirement B, and requirement B is determined to be of higher priority than requirement C, then requirement B should be of higher priority when compared to requirement C. Despite this, AHP lets the decision maker perform the last comparison. Because of this redundancy AHP can indicate inconsistent judgments (such as claiming B to be of higher priority than C in this example) [3].

The redundancy of the comparisons would be unnecessary in case of the decision makers being perfectly consistent. In such a case only  $(n - 1)$  comparisons would be enough to calculate the relative intensity of the remaining comparisons. This implies that the least effort required by a decision maker is to create a minimal spanning tree in a directed graph (i.e. the graph is at least minimally connected). In the directed graph which can be constructed by the comparison provided, there is at least one path between the requirements not pair-wise compared [3].

#### C-1. How to implement Minimal Spanning Tree in Requirement Prioritization:

- (1) As preparation,  $(n - 1)$  unique pairs of requirements are outlined to construct a minimal spanning tree.
- (2) As execution, all outlined pairs of requirements are compared using the scale in Table 1.
- (3) As presentation, the missing intensities of importance are computed by taking the geometric mean of the existing intensities of all possible ways in which they are connected. After that, AHP is used as usually.

The minimal spanning tree approach is supposed to be very fast as it dramatically reduces the number of pair wise comparisons. On the other hand, it is more sensitive to judgmental errors since all redundancy has been removed.

### D. Bubble Sort

Bubble sort is one of the simplest and most basic methods for sorting elements with respect to a criterion [22]. It is also a candidate method for prioritizing software requirements, since the actual prioritizing process can be

viewed as sorting requirements (i.e. the elements) according to their priorities (i.e. the criterion) [3].

Interestingly, bubble sort is closely related to AHP. As with AHP, the required number of pair wise comparisons in bubble sort is  $\frac{n(n-1)}{2}$ . But, the decision maker only has to determine which of the two requirements is of higher priority, not to what extent [3].

#### D-1. How to implement Bubble Sort in Requirement Prioritization:

- (1) As preparation, all requirements are outlined in a vector.
- (2) As execution, all requirements are compared according to bubble sort algorithm.
- (3) As presentation, the sorted vector is outlined. The result of the process is a vector where the original order of the requirements has changed. The least important requirement is at the top of the vector, and the most important requirement is at the bottom of the vector.

The result of a bubble sort is requirements are ranked according to their priority on an ordinal scale.

### E. Binary Search Tree

A binary tree is a tree in which each node has at most two children. A special case of a binary tree is a binary search tree where the nodes are labeled with elements of a set [22]. Consider the elements of the set as the candidate requirements. This is of interest for prioritizing purposes since an important property of a binary search tree is that all requirements stored in the left sub tree of the node  $x$  are all of lower priority than the requirement stored at  $x$ , and all requirements stored in the right sub tree of  $x$  are of higher priority than the requirement stored in  $x$ . If the nodes in a binary search tree are traversed using in order traversing method, then the requirements are listed in sorted order. Consequently creating a binary search tree with requirements representing the elements of a set becomes a method for prioritizing software requirements [3].

Prioritizing  $n$  software requirements using the binary search tree approach involves constructing a binary search tree consisting of  $n$  nodes. The first thing to be done is to create a single node holding one requirement. Then the next requirement is compared to the top node in the binary search tree. If it is of lower priority than the node, it is compared to the node's left child, and so forth. If it is of higher priority than the node, it is compared to the node's right child, and so forth. Finally the requirements are inserted into the proper place and the process continues until all requirements have been inserted into the binary search tree [3].

#### E-1. How to implement Binary Search Tree in Requirement Prioritization:

- (1) As preparation, all candidate requirements are outlined.

- (2) As execution, select the requirements one at a time and a binary search tree is created.
- (3) As presentation, the binary search tree is traversed using in order traversing and nodes are added to a list. The requirements having the lowest priority then come first in the list. Then the list is printed.

Since the average path length from the root to a leaf in a binary search tree is  $O(\log n)$ , inserting a requirement into a binary search tree takes on the average  $O(\log n)$  time. Consequently, inserting all  $n$  requirements into a binary search tree takes on the average  $O(n \log n)$  time. In this case, too, the requirements are ranked on an ordinal scale.

#### F. Priority Groups

In some software development projects, one set of requirements can clearly be of a different kind of importance than another set. One way to reduce the required effort is therefore not to compare the requirements in these distinct sets. Thus another candidate method is to initiate the prioritizing process by dividing the requirements into separate groups based on a rough prioritization. Subsequently, the groups can be internally ranked either by using a suitable approach for ordering the requirement, for example, using AHP or to continue with another grouping of even finer granularity [3].

The primary gain is that, it is not necessary to compare high priority requirements with requirements of low priority, since they are placed in different groups. The actual choice of the number of groups depends on the situation as well as the knowledge of the people performing the prioritization. A simple strategy suggests using three distinct groups: low, medium and high priority. It may even be the case that the high priority requirements must be implemented, and hence there is no need to prioritize between them. In the same way the low-priority requirements may perhaps be postponed to a later release [3].

##### F-1. How to implement Priority Groups in Requirement Prioritization:

- (1) As preparation, all candidate requirements are outlined.
- (2) As execution, each of the requirements is put into one of the three groups. In groups with more than one requirement, three new subgroups are created and the requirements are put into these groups. This process is continued recursively to all groups.
- (3) As presentation, the requirements are printed from left to right.

To guarantee that the correct ordering of the requirements is obtained, it is necessary to ensure that the tail of one group is having higher priority than the head of the following group. This comparison between tail and head in the groups must continue until the requirements are in the correct order. This is one way of minimizing the risk of ending up with the requirements in the wrong order. The priority grouping approach can hence be divided into two

possible approaches: grouping without tail-head comparison and grouping with tail-head comparison [3].

#### G. Planning Game (PG)

In extreme programming the requirements are written down by the customer on a story card which is then divided into three different piles. According to Beck [23], the piles should have the names; “those without which the system will not function”, “those that are less essential but provide significant business value” and “those that would be nice to have”. At the same time, the programmer estimates how long time each requirement would take to implement and then begin to sort the requirements into three different piles, i.e. sort by risk, with the names; “those that can be estimated precisely”, “those that can be estimated reasonably well” and “those that cannot be estimated at all”. Final result of this sorting is a sorted list of requirements on an ordinal scale. Since PG takes one requirement and then decides which pile the requirement belongs to and each requirement is not being compared to any other requirement, the time to prioritize  $n$  requirements is  $n$  comparisons. This means that PG is very flexible and can scale up to rather high numbers of requirements, without taking too long time to prioritize them all.

##### G-1. How to implement Planning Game in Requirement Prioritization:

- (1) As preparation, all candidate requirements are outlined.
- (2) As execution, each requirement is taken and put in the appropriate pile. This process continues until all requirements are sorted.
- (3) As presentation, all requirements in the pile “those without which the system will not function” are considered and system is developed with them.

#### H. 100 Points Method

In this method each stakeholder gets hundred points of some value. With these points they should “purchase ideas”. Each person writes down on a paper how much he/she thinks that one requirement is worth. When all the participants have written down their points, one person calculates, by taking the paper and summing up the points that each requirement has got, and presents the cumulative voting results. The requirement that has got the highest score is the most important requirement.

Theoretically 100P is equally flexible as PG when it comes to the number of comparisons, i.e.  $n$  requirements takes  $n$  comparisons. Hence, it should be a really fast and scalable method, also in comparison to both AHP and BST. However, even though it has the same amount of comparisons as PG, i.e.  $n$ , it probably would take longer time to do the actual comparisons. The reason for this is that, while in PG the decision should be in which pile to put a requirement, i.e. ordinal scale, which is the same scale as BST, in BST the decision should be if requirement  $A$  is more or less important than requirement  $B$ . For 100P the scale is ratio, which is the same scale as for AHP. So the person that

does the prioritization has to consider to which extent one requirement is more or less important than the other. At the same time he/she has only a small amount of points to distribute, which probably also takes some time to take into account in the distribution of points to the different requirements.

#### *H-1. How to implement 100 Points Method in Requirement Prioritization:*

- (1) As preparation, all candidate requirements are outlined.
- (2) As execution, 100 points are distributed among the requirements, according to their importance.
- (3) As presentation, points are summed up for each requirement and requirements sorted according to their total point.

#### *I. Planning Game Combined with AHP (PGcAHP)*

Karlsson et al [24] did a study regarding the difference between PG and AHP where in the discussion section; they stated that it would be interesting to combine PG with AHP, so that this combined method would use the strengths in each method and eliminate their weaknesses.

The strength of PG is that it is rather fast [24], i.e. for  $n$  requirements the total prioritization takes  $n$  comparisons, but the limitation is that it is not possible to calculate the consistency rate and say how important one requirement is against another, only that it is more important. The strength of the AHP is that it is possible to calculate the consistency rate and know exactly how important one requirement is against the other. The limitation of the method is that the number of pair-wise comparisons among the requirements grows exponentially with the increase in the number of requirements. For  $n$  requirements the total number of pair-

wise comparisons is:  $\frac{n(n-1)}{2}$ .

The idea of Karlsson et al [24] is to first divide all requirements into three piles with PG and then the most important pile is taken and requirements within the pile are prioritized using AHP. The advantage of this method is: requirement engineer could pay attention to the most important requirements instead of those less important which saves time. The fewest number of comparisons with PGcAHP would be equal to PG, i.e.  $n$ . However this would indicate that there is at most one requirement that is very important. The other requirements are either important or not important. In that case the idea would be to redo the PG part, or to apply AHP on the middle pile of requirements. The above scenario with maximum one *very important requirement* is only theoretical, logically an application need more than one requirement to be able to function properly.

Now let's take a closer look at the worst and the best case of the method. The best case with the lowest number of comparisons is equal to not putting any requirements in the most important pile, i.e. the number of comparisons is equal

to  $n$ . The worse case is that all the requirements have been placed into the most important pile. That would lead to the longest time to prioritize the requirements, which would be equal to  $n + \frac{n(n-1)}{2}$ , i.e. the number of comparisons for PG + the number of comparisons for AHP.

#### *I-1. How to implement PGcAHP Method in Requirement Prioritization:*

- (1) As preparation, all candidate requirements are outlined.
- (2) As execution, each requirement is taken and put in the appropriate pile. This process continues until all requirements are sorted. Then all requirements within the pile "those without which the system will not function" are put in a matrix and compared using the scale in Table 1 like AHP.
- (3) As presentation, the 'averaging over normalized columns' method (based on the pair wise comparisons) is used to estimate the relative priority of each requirement and consistency ratio of the pair-wise comparisons is calculated using methods provided by AHP. The consistency indicates how trustworthy the results are and also how much judgment error that the analyst has done in the prioritization.

Mathematically it is more suitable to apply PGcAHP, iff, the amount of requirements is more than 21 and if we believe that the number of very important requirements falls within 80% or 90%. If the number of requirements is less than 80% it is better to use AHP. However, this is purely mathematical. In the real life, we do not know how important the requirements are before we prioritize them against each other. Hence, it is not possible to say if the 80% or 90% level is applicable to our problem(s).

## IV. EVALUATION FRAMEWORKS

### *A. Introduction*

The objective is to evaluate the prioritizing methods presented in the previous section to find out which method takes the shortest amount of time in combination to yield the most accurate result and are able to scale up when more requirement are added, from the point of view of prioritizing personnel (personnel could be technical, economical and/or somebody that represent the customer, either by knowledge, behaviors, or any other way that could interesting for the successes for the project) who are going to prioritize the requirements. This section outlines the framework of the evaluation which has been carried out in the form of an experiment. The framework is highly influenced by the experimental approach outlined in [25].

### *B. Preperation*

With the motivation of gaining a better understanding of requirement prioritization, we performed a single project study [25] with the aim of characterizing and evaluating the

requirement prioritizing methods under observation from the perspective of users and project managers. This experiment was carried out by 10 persons consist of software developers, project managers, faculty members of computer science and persons without computer science background. They were asked to prioritize 14 features of a mobile set using the prioritization techniques under observation. The requirements were prioritized by the participants independently, and to the best of their knowledge. The quality requirements were prioritized without taking the cost of achieving the requirements into account. That is, only the importance for the customers was considered. Moreover, the requirements were considered orthogonally, i.e. the importance of one requirement is not interdependent on another [3].

Only one method was studied each day to minimize the influence of the order of the methods, and to reduce the influence of the persons remembering the priorities of the requirements using the previous methods. Each day, 15 minutes were allocated for presenting the method which was under observation on that day and after getting the confirmation from each participant that the method was understood clearly, 60 minutes were allocated for completion of the experiment of that day. Each participant was supplied with necessary papers and time taken by each participant to complete the experiment was recorded separately.

### C. Threats to Validity

When reading a result from an experiment, one of the most important questions is: How valid the result is? That makes validity of the result an important question to consider when an experiment is designed [26]. The overall objective of the experiment was evaluation of some requirement prioritization techniques by making some comparisons among them. We do not argue that the results obtained in this evaluation can be generalized and used by any user in any environment for any application. Rather, we tried to illustrate the requirement prioritizing methods to gain a better understanding of them. The following threats have been identified:

#### C-1. Few persons involved in the experiment:

The significance of the results is limited due to involvement of few persons (10 persons) with the experiment. That's why the outcomes were more inconclusive, and hence can be regarded as a partial threat to the evaluation. However, if requests to attend to the experiment are going to a large population, there is a greater chance that the risk would be minimized.

#### C-2. Too few requirements:

In the analysis of the data, it became obvious that the experiment had too few requirements. However, before the experiment it was discussed whether it would be possible to consider more than 14 requirements, but since there was a time limit, i.e. how much time the participants could participate, the number of requirements had to be limited. To

really reflect a real project, the number of requirements should be a couple of hundred; this would be more or less impractical to handle within the limited timeframe of this experiment, therefore the decision was taken that the number of requirements should only be 14.

#### C-3. Hypothesis guessing:

There was a risk that the participants might try to figure out what the intention and the purpose of the experiment were, and then they would answer to satisfy this intention and purpose. That would result in a misleading conclusion.

#### C-4. Requirements are interdependent:

In practice, the interdependence between the requirements must be considered. None of the prioritizing methods described in this paper provides means for handling interdependence; hence this limitation of the experiment is not believed to influence the actual evaluation of the different methods.

#### C-5. Only non functional requirements considered:

This experiment was only concerned with non functional requirements. However, we don't think it to be a major threat to the results from the experiment.

#### C-6. Offline evaluation:

The evaluation was carried out independently from a software project which may be considered as a potential problem for this experiment. However, it is not regarded as being a major threat as the main objective of this evaluation was to gain understanding and illustrate a number of potential methods for prioritizing software requirements.

It is always important to identify threats in an experiment in order to allow for determining both the internal and external validity of the results attained. Thus, the above potential threats should be kept in mind when analyzing the results [3].

### D. Analysis of Collected Data

The testing begins with the first question of every method; followed by the second and third and so on. For each question, participants ranked each method and finally mean value was taken. An important notice is that all the numbers have been rounded to two significant digits.

#### D-1. Ease of Use:

The first question that the participants were asked was how easy they thought that the method was. The result of the question is shown in Figure 1.

Figure 1 clearly indicates that Planning Game (PG) is the easiest method and AHP is the toughest one.

TABLE IV. Comparison among the methods f

Evaluation Criteria	AHP	Hierarchy AHP	Minimal Spanning Tree	Bubble Sort	Binary Search Tree
Consistency (Yes / No)	Yes	Yes	No	No	Yes
Scale of Measurement	Ratio	Ratio	Ratio	Ordinal	Ordinal

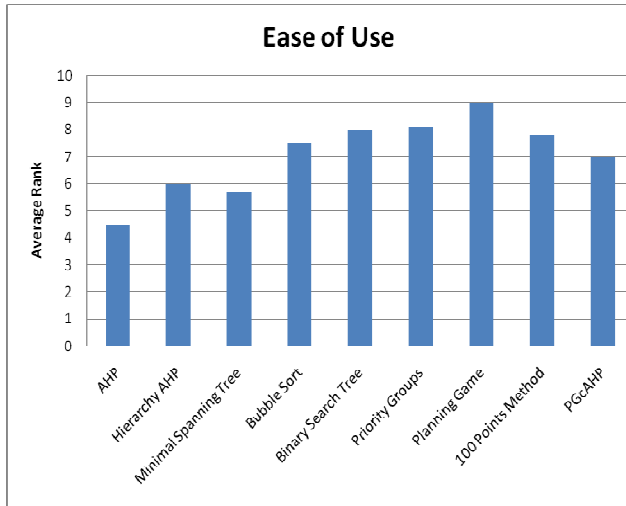


Figure 1: Comparison among the methods for the criteria "Ease of Use"

#### D-2. Certainty:

The second question that the participants were asked was how certain they were about the end result obtained from the methods under consideration. The result is presented in the Table 4.

Though no statistical data was collected for this criterion, however, most participants think that AHP, Hierarchy AHP, Binary Search Tree, Planning Game (PG) and PGcAHP produce consistent result.

#### D-3. Total Time Taken:

Next question was how long time it took for the participants to perform the prioritization with the method under consideration. The result is presented in Figure 2.

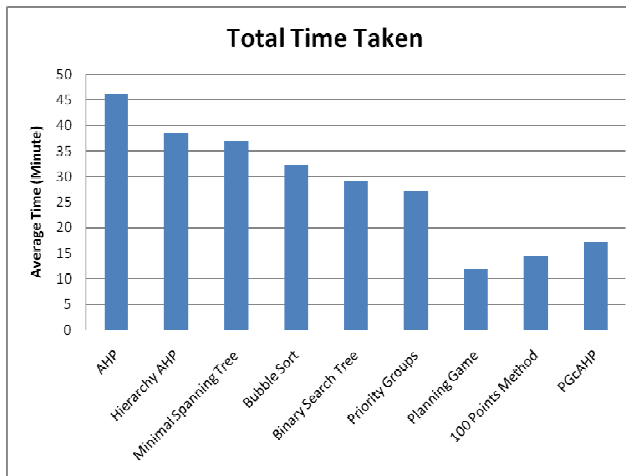


Figure 2: Comparison among the methods for the criteria "Total Time Taken"

The above graph clearly indicates that Planning Game (PG) is the fastest and AHP is the slowest among the prioritization techniques under consideration.

#### D-4. Scalability:

In next stage participants were asked to arrange the methods according to how they believe that the methods would work with many more requirements than the 14 considered in the experiment. The result is presented in Figure 3.

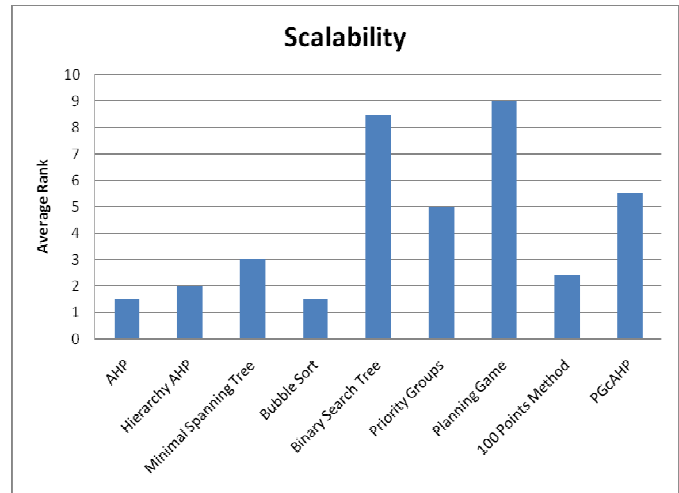


Figure 3: Comparison among the methods for the criteria "Scalability"

This graph indicates that most of the participant think that PG and BST are more scalable than other methods where as AHP, Hierarchy AHP and Bubble Sort are not suitable for large number of requirements.

#### D-5. Accuracy:

In this stage the participants were asked to arrange the methods under consideration according to their opinion about accuracy of the result produced by each method. However, there was a minor error that was overlooked in the experiment and it was that the words "accuracy" and "certainty" were used as meaning the same thing. Hence, final accuracy was compared with certainty in section D-2. The result is shown in figure 4.

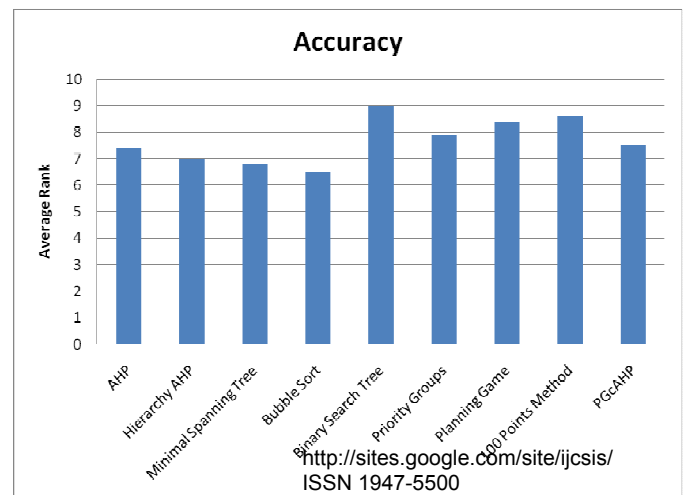


Figure 4: Comparison among the methods for the criteria "Accuracy"



From figure 4 in can be imagined that BST and 100 Points Method yield the best result. It was expected that AHP would produce the most accurate result as in this method requirements were prioritized according to mathematical rules. An explanation to why AHP more or less did so poorly here could be that the participants did not understand how to read out the matrix that presented the prioritization results.

#### D-6. Total Number of Comparisons:

The participants were asked to keep record of how many comparisons were required for each method. The result is shown in Figure 5.

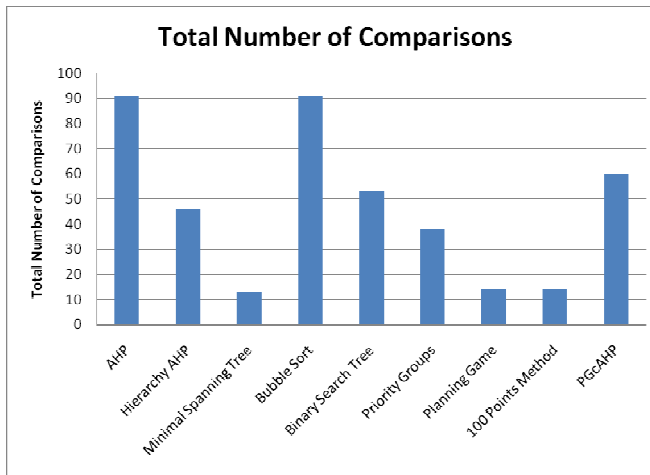


Figure 5: Comparison among the methods for the criteria "Total Number of Comparisons"

Figure 5 shows that Minimal Spanning Tree requires lowest number of comparisons where as AHP and Bubble Sort require highest number of comparisons.

#### V. FINDING THE BEST TECHNIQUE

After collecting data based on above motioned criteria, we assigned weight for each criterion and they applied a formula (8), (9) and (10) to find out the best technique for requirement prioritization. Each of the evaluation criteria was assigned weight according to Table 5.

TABLE V. Weight Table for Each Criterion

Name of the Criterion	Weight
Ease of Use	10
Certainty	3
Total Time Taken	9
Scalability	7
Accuracy	8.5
Total Number of Comparisons	8

Then following formulae were used to calculate overall score by each of the prioritization techniques under consideration.

$$C_{ij} = W(C_i) * ((N + 1) - Rc_i(M_j)) \text{ [Except } C_2]$$

(8)

$$C_{2j} = W(C_2) * ((N+1) * IsCertain(M_j))$$

(9)

where IsCertain = 1 if  $M_j$  has Certainty else IsCertain = 0

$$OS(M_j) = \frac{\sum_{i=1}^{NC} C_{ij}}{NC}$$

(10)

Here,

N = Number of Techniques used

NC = Number of Criteria

$C_{ij}$  = Score of Technique j in Criteria i

$C_2$  = Certainty

$W(C_i)$  = Weight of  $C_i$

$Rc_i(M_j)$  = Ranking of Technique j in Criteria i

$OS(M_j)$  = Overall Score of Technique j

The result after calculation is shown in Figure 6.

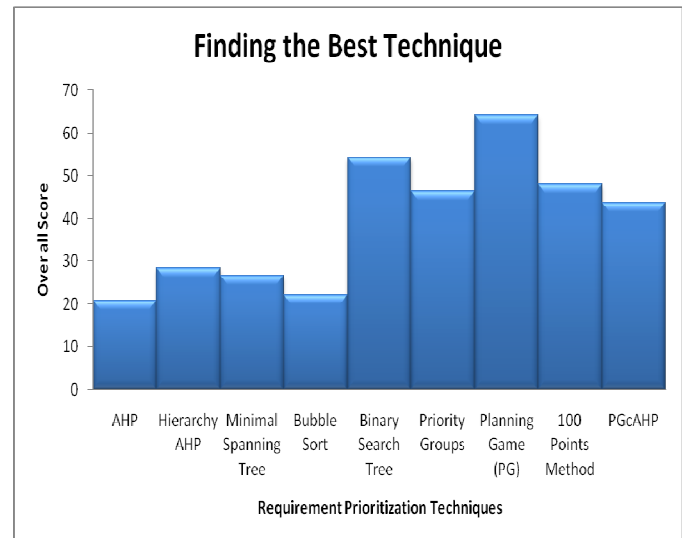


Figure 6: Comparison among the methods on the basis of weighted value

Figure 6 clearly indicates that among all the requirement prioritization techniques under consideration, Planning Game (PG) is supposed to be the best one on the basis of the mentioned evaluation criteria.

This order of the requirement prioritization techniques obtained from this experiment, however, is not a global one as rankings can be reordered if criterion weights are assigned differently. Nevertheless, the technique and formulae used here to compare among different prioritization methods can be used in any scenario with appropriate criterion weights suitable for that scenario.

## VI. DISCUSSION

Final outcome of the experiment says that Planning Game (PG) is supposed to be the best method for prioritizing software requirements. It is an easy method which produces one of the most accurate results and it is rather easy to handle even if there are many more requirements. The worst candidate according to the result is Analytical Hierarchy Process (AHP) as according to the participants it is the toughest method and it is difficult to scale up for a large number of requirements. Although the results have indicates that PG is the best and AHP is the worst among the candidate techniques of this experiment, there are some variables that may have some impact on the result.

- (1) Most of the participants had previous experience and knowledge about the working mechanism of Planning Game (PG) technique. Hence, they found it comfortable while prioritizing requirements.
- (2) One the other hand, some of the participant had no previous experience about some methods such as AHP, Hierarchy AHP and hence they may not find it convenient to apply these methods to prioritize requirement.

However, these are some minor issues having impact on the obtained result, and it would be interesting to evaluate these issues in various experiments in future.

## VII. CONCLUSION

Methods for establishing priorities are of great importance in software development, since the developers' best effort can more easily be focused on the issues which matter most for the success of the system [3]. Therefore we have evaluated and characterized nine different methods for establishing priorities. In our evaluation we found PG to be the most promising approach as it yields one of the most trustworthy results by taking least time and it also works fine when the number of requirements increases.

Of course, the other methods considered in this experiment could be combined in many different ways to provide a more efficient approach for prioritizing requirements where the required number of comparisons could be reduced in an efficient manner.

Interestingly, after this experiment, we found that evaluating the requirement prioritization techniques in group sessions to be a means of communicating knowledge, achieving consensus and effective for identifying potential problems in the requirements. Using a group rather than individuals in this regard, helps the participants to bring out their particular knowledge on which they judge the requirements and that's why, as soon as the results were made visible to the participants, they immediately felt that the results reflected their judgments.

## REFERENCES

- [1] Sawyer P (2000) Packaged software: challenges for RE. Proc Int Workshop on Req Eng: Foundations of Software Quality. Stockholm, Sweden, pp 137-142
- [2] J. Karlsson, K. Ryan, Prioritizing requirements using a cost-value approach, IEEE Software 14 (5) (1997) 67-74.
- [3] Karlsson, J., Wohlin, C., & Regnell, B. (1998). An evaluation of methods for prioritizing software requirements. Information and Software Technology, 39(14-15), 939-947.
- [4] Hatton, S. (2007). Early prioritisation of goals. In Advances in conceptual modeling – Foundations and applications (pp. 235-244).
- [5] Ngo-The, A., & Ruhe, G. (2005). Decision support in requirements engineering. In A. Aurum & C. Wohlin (Eds.), Engineering and managing software requirements (pp. 267-286): Springer Berlin Heidelberg.
- [6] H. In, D. Olson, and T. Rodgers. Multi-Criteria Preference Analysis for Systematic Requirements Negotiation. In 26th Annual International Computer Software and Applications Conference, Oxford, England, August 2002.
- [7] J. Karlsson, Software requirements prioritizing, in: Proc. of 2nd IEEE Int. Conf. on Requirements Eng. (1996) pp. 110-116.
- [8] F.Moisiadis. The fundamentals of prioritising requirements. In System Engineering, Test and Evaluation Conference, Sydney, Australia, 2002.
- [9] A. Ngo-The and G. Ruhe. Requirements Negotiation under Incompleteness and Uncertainty. In Software Engineering Knowledge Engineering 2003 (SEKE 2003), San Francisco, CA, USA, July 2003.
- [10] A. Sivzattian and B. Nuseibeh. Linking the selection of requirements to market value: A portfolio-based approach. In REFS 2001, 2001.
- [11] J. Siddiqi, M.C. Shekaran, Requirements engineering: the emerging wisdom, IEEE Software 13 (2) (1996) 15-19.
- [12] M. Lubars, C. Potts, C. Richter, A review of the state of the practice in requirements modeling, in: Proc. of the IEEE Int. Symp. on Requirements Eng. (1993) pp. 2-14.
- [13] T.L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, Inc. (1980).
- [14] G.R. Finnie, G.E. Wittig, D.I. Petkov, Prioritizing software development productivity factors using the analytic hierarchy process, J. Systems Software 22 (2) (1993) 129-139.
- [15] C. Douligeris, I.J. Pereira, A telecommunications quality study using the analytic hierarchy process, IEEE J. Selected Areas Commun. 12(2) (1994) 241-250.
- [16] J. Karlsson, Towards a strategy for software requirements selection. Licentiate thesis 513, Department of Computer and Information Science, Linköping University, 1995.
- [17] Paetsch F., Eberlein A., Maurer F., "Requirements Engineering and Agile Software Development", Proceedings of the 12 IEEE International workshop, IEEE Computer society, 2003, pp 1-6
- [18] Regnell B., Höst M., Natt och Dag J., Beremark P., Hjelm T., "An Industrial Case Study on Distributed Prioritisation in Market-Driven Requirements Engineering for Packaged Software", Requirements Engineering, Feb 2001, vol. 6, number 1, pp 51-62
- [19] Saaty T. L., "That Is Not the Analytic Hierarchy Process: What the AHP Is and What It Is Not", Journal of Multi-Criteria Decision Analysis, 1997, vol. 6, No: 6, pp 324 - 335
- [20] Gass S., Rapcsák T., " Singular value decomposition in AHP", European Journal of Operational Research, 2004, Vol. 2004, No. 3, pp 573-584
- [21] A. Davis, Software Requirements: Objects, Functions and States. Prentice-Hall International, Englewood Cliffs, New Jersey, 1993.
- [22] A.V. Aho, J.E. Hopcroft, J.D. Ullman, Data Structures and Algorithms. Addison-Wesley, Reading, MA, 1983.

- [23] Beck K., "Extreme programming: explained", Addison-Wesley, USA, December 2001, 7th edition.
- [24] Karlsson L., Berander P., Regnell B., Wohlin C., "Requirements Prioritisation: An Exhaustive Pair-Wise Comparison versus PG Partitioning", EASE '04- Empirical Assessment in Software Engineering, 2004
- [25] V.R. Basili, R.W. Selby, D.H. Hutchens, Experimentation in software engineering, IEEE Trans. Soft. Eng. 12 (7) (1986) 733-743.
- [26] Wohlin C., Runeson P., Höst M., Ohlsson M., Regnell B., Wesslén A., "Experimentation in software engineering: An introduction", Kluwer Academic Publishers, USA, 2000

The Standish Group International Inc, "CHAOS Summary 2009: The 10 Laws of CHAOS", found at [http://www.statelibrary.state.pa.us/portal/server.pt/document/690719/chaos\\_summary\\_2009\\_pdf](http://www.statelibrary.state.pa.us/portal/server.pt/document/690719/chaos_summary_2009_pdf)

#### AUTHORS' PROFILE

Mohammad Shabbir Hasan received his B.Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2008. His research interest includes different areas of Software Engineering like Requirement Engineering, Software Metric, Software Security and Software Maintenance. He has coauthored numerous research papers published in International Journals and Conference Proceedings. Currently he is working as a researcher of Panacea Research Lab, Dhaka, Bangladesh. He is also a lecturer of Department of Computer Science and Engineering in Institute of Science, Trade and Technology, Dhaka, Bangladesh.



Abdullah Al Mahmood received his B.Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2009. His research interest includes Robotics, Artificial Intelligence, Internet Security and various areas of Software Engineering like Requirement Gathering, Requirement Prioritization and Software Security. He has coauthored a good number of research papers published in International Journals and Conference Proceedings. Currently he is working as a researcher of Panacea Research Lab, Dhaka, Bangladesh. He is also a lecturer at Institute of Science, Trade and Technology, Dhaka, Bangladesh and a Project Coordinator of Technocrats BD.



Md. Jahangir Alam received his B.Sc. (Engg.) and M.Sc. in Computer Science and Engineering from Dhaka International University (DIU), Bangladesh. His research interests include Software Engineering and Data Structure. Currently he is working as a Senior Lecturer of Department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh.



Sk. Md. Nahid Hasan received his B.Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2010. His research interest includes different areas of Software Engineering like Software Metric, Requirement Gathering, Software Security and Software Maintenance. Digital Image Processing is also one of his research concerns. Currently he is working as a researcher of Panacea Research Lab, Dhaka, Bangladesh. He is also a lecturer of Department of Computer Science and Engineering in Institute of Science, Trade and Technology, Dhaka, Bangladesh.



Farin Rahman received her B. Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2009. Her research interest encompasses with different sectors of Software Engineering like Requirement Engineering, Software Security and Software Maintenance. She is working as a researcher in Panacea Research Lab and also as a lecturer in Daffodil Institute of Information Technology, Dhaka, Bangladesh.



## Routing Fairness in Mobile Ad-Hoc Networks : Analysis and Enhancement

MANIYAR SHIRAZ AHMED<sup>1</sup>, HAJI MOINUDDIN<sup>2</sup>, FAZEELATUNNISA<sup>3</sup>

<sup>1</sup>Lecturer, Department of Computer Science & Information Systems,  
Najran University, Najran, Saudi Arabia,  
samaniyar@nu.edu.sa, shiruraz@yahoo.com

<sup>2</sup> Lecturer, Department of Computer Science & Information Systems,  
Najran University, Najran, Saudi Arabia  
haji.moinuddin@yahoo.com, hmaddin@nu.edu.sa  
<http://www.hajimoinuddin.co.nr/>

<sup>3</sup> Lecturer, Department of Computer Science & Information Systems,  
Najran University, Najran, Saudi Arabia  
ftshiraz@nu.edu.sa, fazeel\_ab@hotmail.com

**Abstract:** With the rapid advances in wireless and semiconductor technologies mobile connectivity became cheap and ubiquitous. One of the major challenges facing Mobile Ad-Hoc Networks (also known as MANETs) is the absence of a proper routing protocol that provides good fairness and scalability, low overhead, low end-to-end delays, seamless connectivity and good quality of service. This paper studies the fairness of routing protocols for MANETS. In this paper we propose routing segments methods to solve the problem of lack of fairness in routing.

**Keywords:** MANETS, Fairness, Segments, Scalability

### I. INTRODUCTION

Wireless devices are becoming ubiquitous, with the ever increasing advances in wireless and mobile computing. Improved protocols must be developed to support these new mobile devices/ MANETS[3] and to see that these devices do not overload the existing infrastructure network. The effort in this endeavor is to provide anytime, anywhere connectivity for unlimited mobile devices without overloading the associated infrastructure networks.

Most protocols in place suffer from low quality of service and overload the network with a large percentage of overhead (control data) when compared to the data packets. Any improvement in the routing protocol should be an extendable architecture to support high number of mobile units and at the same time ensures a good quality of service.

Mobile routing protocols have been attracting the attention of a major section of the research community as is evident from the large number of ongoing projects at various universities and institutions on this topic. Numerous architectures have been proposed earlier

### II. RELATED WORK

Routing protocols form the heart of any MANET, which have not evolved as much to support a large amount of mobile units. The performance of most routing protocols degrades with the increase in mobile nodes,[9] leading to higher end-to-end delay, more dropped packets and low quality of service (QoS).

Dynamic routing protocols can be classified in several ways. Basically it is classified into two (a) exterior protocols versus interior protocols, and (b) distance-vector versus

link-state protocols.[4] The first classification is based on where a protocol is intended to be used: between your network and another's network, or within your network. The second classification has to do with the kind of information the protocol carries and the way each router makes its decision about how to fill in its routing table.

#### **(a) Exterior vs. Interior Protocols**

Dynamic routing protocols are generally classified as an exterior gateway [1] protocol (EGP[2]) or an interior gateway protocol (IGP). An exterior protocol carries routing information between two independent administrative entities, such as two corporations or two universities. Each of these entities maintains an independent network infrastructure and uses an EGP to communicate routing information to the other. Today, the most common exterior protocol is the Border Gateway Protocol (BGP). It is the primary exterior protocol used between networks connected to the Internet, and was designed specifically for such purposes.

In contrast, an interior protocol is used within a single administrative domain, or among closely cooperating groups. In contrast to the exterior protocols, IGPs tend to be simpler and to require less overhead in a router. Their primary drawback is that they cannot scale to extremely large networks. The most common interior protocols in IP networks are the Routing Information Protocol (RIP), Open Shortest Path First (OSPF), and the Enhanced Interior Gateway Routing Protocol (EIGRP). The first two are open standards adopted or developed by the Internet community, while the third is a proprietary protocol designed by Cisco Systems for use on their routers.

#### **(b) Distance-Vector vs. Link-State Protocols**

Another way to classify dynamic routing protocols is by what the routers tell each other,

and how they use the information to form their routing tables. Most protocols fit into one of two categories.

The first of these categories is distance-vector protocols. In a distance-vector protocol, a router periodically sends all of its neighbor's two pieces of information about the destinations it knows how to reach. First, the router tells its neighbors how far away it thinks the destination is; second, it tells its neighbors what direction (or vector) to use to get to the destination. This direction indicates the next hop that a listener should use to reach the destination, and typically takes the form "send it to me, I know how to get there." For example, RIP route updates simply list a set of destinations that the announcing router knows how to reach, and how far away it thinks each destination is. The receiver infers that the next hop to use is the announcing router. However, an update can also take the form "send it to this other router who knows how to get there." This second form is usually used only when the router that should be used to reach the destination cannot (or will not) speak the routing protocol being used by the other routers. Not all routing protocols support this form of third-party route update.

The other part of the protocol, the distance, is where distance-vector protocols differ. In each case, the protocol uses some metric to tell the receiving routers how far away the destination is. This metric may be a true attempt at measuring distance (perhaps using a periodic measure of the round trip time to the destination), something that approximates distance (such as hop count), or it may not measure distance at all. Instead, it may attempt to measure the cost of the path to the destination. It may even involve a complex computation that takes into account factors like network load, link bandwidth, link delay, or any other measure of the desirability of a route. Finally, it may include an administrative weight that is set by a network administrator to try to cause one path to be preferred over another.



In any case, the metric allows a router that hears about a destination from multiple routers to select the best path by comparing the "distance" of the various alternatives. How the comparison is made depends heavily upon how metric is computed. For example, the metric in RIP route updates is defined to be a hop count, in which one hop is supposed to represent handling by one router. A destination with a hop count of 16 is considered unreachable. When a router receives RIP updates from different routers referring to the same destination network, it selects the router that is announcing the lowest metric. If this metric is lower than the metric for the route that is currently in its routing table, the router replaces its routing table entry with the new information from the other router.

In contrast, in a link-state protocol, a router does not provide information about destinations it knows how to reach. Instead, it provides information about the topology of the network in its immediate vicinity. This information consists of a list of the network segments, or links, to which it is attached, and the state of those links (functioning or not functioning). This information is then flooded throughout the network. By flooding the information throughout the network, every router can build its own picture of the current state of all of the links in the network. Because every router sees the same information, all of these pictures should be the same. From this picture, each router computes its best path to all destinations, and populates its routing table with this information. How a router determines which path is best is up to each protocol. In the simplest case, a router may simply compute the path with the least number of hops. In a more complex protocol, the link-state information may include additional information to help a router determine the best path. Such information may again include the bandwidth of the link, the current load on the link, administrative weights, or even policy information restricting which packets may

traverse the link. For example, a given link might not be allowed to carry confidential information.

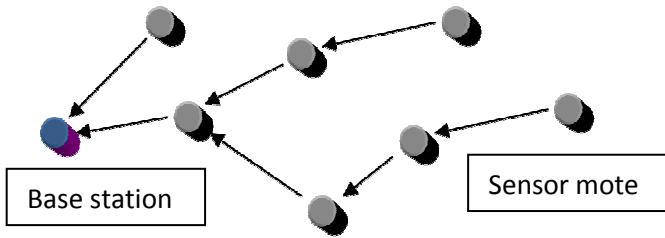
Distance-vector and link-state protocols have their own strengths and weaknesses. In a properly functioning and configured network, either type yields a correct determination of the best path between any two points.

### III. PROPOSED PROTOCOL

Binding refers to keeping the network together, issuing routing updates, keeping track of nodes entering and exiting the network etc. As the size of the MANET increases, the control traffic also increases. When nodes are tasked with binding the network as well as data transfer, bottlenecks are created within the network leading not only to battery drain out but slow network performance and unfairness in routing. Hence it is critically important to disassociate both of these functionalities to prevent node failures due to bottle necks and also unfairness and low power conditions. We also have to solve the problem of scalability.

This can be done by managing the manets based on the routing and must implement routing fairness in order to prevent early partitioning of the ad hoc network into disjoint network segments.

Routing fairness [5] in MANETS is essential as it discourages large volume of disjoint network segments. Each sector can have two motes (sensor mote and base and Sensor motes gather data and send to central mote(base station). Motes too far from base station requires intermediate motes to relay, or route, data. Routing structure formed is a tree, rooted at the base station.

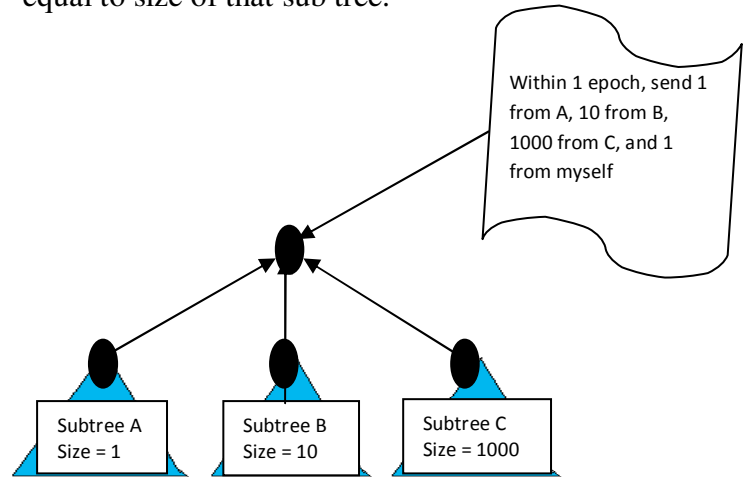


The problems with this kind of routing structures can be like motes closer to the base station has to transmit packets generated locally as well as those generated by downstream motes, these motes likely to become bottlenecks in the system which results in more packets originating further away being dropped (unfairness) loss of packets due to queue overflow and interference during transmission (congestion. unfairness may result in network not retrieving sufficient data from faraway motes to meet application requirements. congestion wastes scarce energy resources.

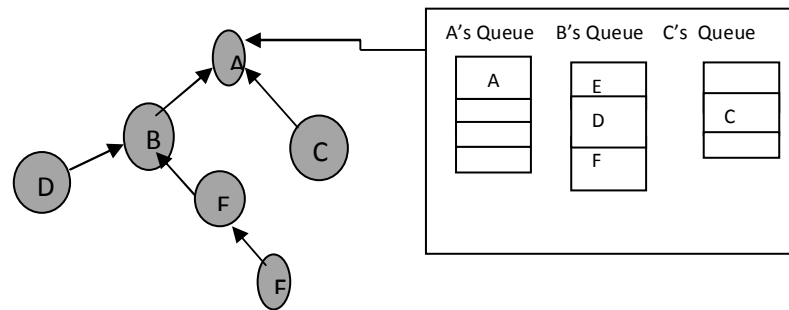
The problem of packets being dropped (unfairness) and be solved by determining maximum application data generation rate and by implementing hop-by-hop Automatic Repeat Request (ARQ).[7] since motes generate data at a rate network can handle, congestion (queue overflow) should not occur. ARQ ensures all packets ultimately reach the base station. BUT difficult to obtain maximum rate for every network configuration underestimation of generation rate reduces effective bandwidth.

#### IV. ROUTING FAIRNESS

To be fair in routing, at base station, same number of packets should receive from each mote. Within each period of time (or epoch), transmit number of packets from each sub tree equal to size of that sub tree.



For this we require per child queue (does not depend on size of sub tree, so can be small and constant), FIFO queues, sub tree size (obtained as before). Then we should check for proof of correctness (by induction).



## V. ENHANCEMENT OF FAIRNESS

The proxies alleviate the unfair advantage that shorter connections have over longer connections. For this we can split longer connections into shorter segments. The throughput of longer connections, however, cannot equal that of shorter connections due to interactions between segments. The packets cannot be sent and received at the proxy at the same time, so the adjacent TCP[6] segments have to transport data in stages

We can improve fairness by having Multiple TCP connections with varying lengths in terms of hop count. Because the Longer connections achieve lower throughput than shorter ones, we have to introduce proxies[8] which improves throughput. For a connection of length 16 hops, the throughput improves from around 22 Kbps to 27 Kbps. Thus there will be improvement in fairness

## VI. ROUTING

Routes can be of two types; first spanning just longer segment, and the second, spanning shorter segments. In the case of routes spanning longer segments, the entire route is divided into shorter segments. This active route is hierarchically managed using two routing protocols; one at the inter-segment level and the other at the intra-segment level.

The entire route from the source to the destination has nodes involving multiple segments divided into shorter segments.

A segment is the route between two gateway nodes or the route between the gateway node and the source node or the destination node. In other words, a route is a connection of one or more segments and segment has a segment-head.

## VII. PROTOCOL IMPLEMENTATION

If data needs to be transferred from one end of the MANET to the other, the source sends a request to its adjacent TCP segment. This TCP segment then forwards the request to the respective TCP segment. The TCP segment which to participate in the data transfer reply back with node addresses of nodes that are active and willing to participate in the route.

Propagation of the route establishment request is between the TCP segments only, which decide on the basis of the instantaneous information able to them. By limiting the propagation of the route request to the segments only the traffic is greatly reduced, because as seen in other protocol the request keeps on propagating due to retransmission from nodes throughout the network until the TTL of the request has expired causing considerable traffic.

Before the source transmits data, it must setup proper segments to be used by the respective protocols. Firstly, depending upon the routing information and number of proxies received for its route, this segment is utilized by the intersegment DSR protocol. Next the source adds intra-segment routing information for the packet to reach the first gateway node.

By adding segments and proxies in this fashion two purposes are served. First CGRS gets the route to the nearest gateway. Second the DSR protocol gets the next hop to inter-segment gateway node. The inter-segment header gets reduced with each hop where as the intra-segment is renewed at each gateway node. A new inter-segment is appended while entering a new segment. Inter-segment routing (DSR) occurs at the gateway nodes while the intersegment routing occurs at both gateway nodes and segment nodes until the data packet reaches the destination.

## VIII. CONCLUSION

This paper studies the dynamic routing fairness for mobile adhoc networks and describes the different existing dynamic routing protocols. Different from existing works, this work considers the routing segments to improve fairness in the routing. We show our assumptions that can be implemented to get more fairness in routing. We proposed routing segment method. The problems in routing fairness and proposed solutions have been discussed.

## REFERENCES

1. Scott M. Ballew. Managing IP Networks with Cisco Routers. Chapter 5 footnote-1. O'Reilly Media. 1997.
2. Scott M. Ballew. Managing IP Networks with Cisco Routers. Chapter 5 footnote-2. O'Reilly Media. 1997.
3. Sarosh Patel, Khaled Elleithy, and Syed S. Rizvi. "Hierarchically Segmented Routing (HSR) Protocol for MANET", 6th International Conference on Information Technology: New Generations ITNG, USA, April 2009.
4. Network Academy . Introduction to Dynamic Routing Protocols, CCNA Exploration Course Booklet: Routing Protocols and Concepts, Chapter-3.
5. Mikael Johansson. Scheduling, routing and power allocation for fairness in wireless networks. In ieev vehicular technology conference (2004)
6. Christopher Ellis. IPv6 Mobile Ad Hoc (MANETs) and Sensor Networks. United States IPv6 Summit 2004
7. Cheng Tien Ee. Congestion control and fairness for many-to-one routing in sensor networks in ACM SenSys
8. Eric Law, UCR. Transport Layer for Mobile Ad Hoc Networks. University of California, March 2005
9. Wikipedia:[http://en.wikipedia.org/wiki/Routing\\_protocols](http://en.wikipedia.org/wiki/Routing_protocols)

## Nano-particle Characterization Using a Fast Hybrid Clustering Technique for TEM Images

**M.A. Abdou**

Researcher, IRI  
Mubarak City for Scientific  
Research and Technology  
Applications  
Alexandria, Egypt.  
[m.abdou@pua.edu.eg](mailto:m.abdou@pua.edu.eg)

**Bayumy B.A. Youssef**

Researcher, IRI  
Mubarak City for Scientific Research  
and Technology Applications  
Alexandria, Egypt.  
[bbayumy@yahoo.com](mailto:bbayumy@yahoo.com)

**W.M. Sheta**

Researcher, IRI  
Mubarak City for Scientific Research  
and Technology Applications  
Alexandria, Egypt  
[wsheta@mucsat.sci.eg](mailto:wsheta@mucsat.sci.eg)

**Abstract-** This Paper introduces a new fast Transmission Electron Microscopy (TEM) images clustering technique. Since analysis of particle sizes and shapes from two-dimensional TEM images is affected by variations in image contrast between adjacent particles, automatic methods requires more efforts. The proposed hybrid method consists of two main steps: automatic segmentation and nano-particles counting. The segmentation procedure begins with an automatic threshold generator and moves towards a high efficient multiple- regions segmentation technique. Results are observed, compared with existing methods and manual counting.

**Keywords:** TEM, Image segmentation, Threshold generator, Nano-particle counting

### 1. INTRODUCTION

(TEM) images are widely used in field of nano-particle characterization. These images should be processed via a clustering technique to obtain the distribution of nano-particles on certain surface. Mean diameter can be measured either manually using a ruler or automatically through computer algorithm. Counting pixels belonging to every cluster, calculation of mean diameter are of great importance. Manual methods require extremely hard work and suffer leak of accuracy. Automatic methods, if used properly, will be easier and attain a mass production in this field. Many researches have been achieved concerning the TEM image analysis. Hideyuki et al. [1] have presented a study of construction a 3D geometric model of dermatan sulfate glycosaminoglycans and collagen fibrils, and to use the model to interpret TEM measurements of the spatial orientation and length of dermatan sulfate glycosaminoglycans in the medial collateral ligament of the human knee. This study shown how a 3D geometric model can be used to provide a priori information for interpretation of geometric measurements from 2D micrographs. Schaeublin et al. [2] presented the usage of TEM image simulations for couple the results from molecular dynamics simulations to experimental TEM images. David et al. [3] synthesized discrete single-element semiconductor nano-wires and multicomposition nano-wire hetero-structures, and then characterized their structure and composition using high-resolution electron microscopy and analytical electron

microscopy techniques. Chinthaka et al. [4] have made a study where the nano-sized fluorapatite particles were synthesized using a precipitation method and the material was characterized using X-ray diffraction and transmission electron microscopy (TEM). Yeng-Ming et al. [5] studied the effect of hydrophobic molecules on the morphology of aqueous solutions of amphiphilic block copolymer, which has potential drug delivery applications. Using cryogenic TEM observations, micelles can clearly be visualized and their core size measured. Kurt et al. [6] demonstrated that TEM techniques are focusing on the determination of parameters, such as shape and size of islands. A successful image contrast analysis in terms of shape and strain demands the application of image simulation techniques based on the many-beam dynamical theory and on structure models refined by molecular dynamics or molecular static energy minimization. Kenta et al. [7] developed a spherical aberration corrected TEM technique that allowed them to obtain clearer images in real space than ever before. They applied this technique to titanium oxide, in which light elements such as oxygen are difficult to observe using TEM because of its small cross section and electronic damage. Wang et al. [8] examined the mean diameter of the PtRu nanoparticles using TEM. Jae et al. [9] investigated the catalysts by employing various physicochemical analyses: X-ray diffraction, TEM and extended X-ray absorption fine structure to investigate the structural modification, and X-ray photoelectron spectroscopy and X-ray absorption-near-edge spectroscopy to characterize the change in electronic features. The data processing was performed with XPSPEAK software program. Because of the lack of Tem image processing, the MRI and Gel image processing were considered in survey. Because both MRI and Gel images are look like TEM images, whereas they have grey level color and the data exists in cluster spots within image background. Many researches have been achieved in the Tem image processing (clustering through image segmentation). Atkins et al. [10] used thresholding and morphology techniques, combined with an anisotropic diffusion process to localize and segment the brain. Ravinda et al. [11] proposed a similar approach. Hahn and Peitgen. [12] proposed a solely intensity-based watershed algorithm, which makes use of a simple merging criterion to avoid the over segmentation problem. Kapur et al. [12] proposed a hybrid approach that



used morphological operations and active contour segmentation. Shattuck et al. [14] used adaptive anisotropic diffusion, edge detection and morphological erosions to identify the brain component. Xu et al. [15] introduced the deformation of the active surface under a gradient vector field computed from a binary edge map. Zeng et al. [15] used a coupled surface evolution to extract bounding surfaces of the cortex. Kim et al. [17] proposed a hierarchical segmentation based on thresholding and the detection of watersheds. They first pre-processed the images to remove noise and enhance contrast, and then thresholding was applied. Takahashi et al. [18] achieved image enhancement and smoothing based on the definition of threshold values, before defining local maxima in order to label the spots. Umer et al. [19], presented a technique that uses the clustering techniques like K-mean and fuzzy C-mean to distinguish between different types of protein spots and unwanted artifacts. Christopher et al. [20] presented a new technique using the labeling of each image pixel as either a spot or non-spot and used a Markov Random Field model and simulated annealing for inference. Neighboring spot labels were then connected to form spot regions. Feature extraction is usually based on computing commonly used features: mean variance, coefficient of correlation, contrast, homogeneity, skew, and kurtosis [21]. Texture segmentation has been improved by the use of co-occurrence matrices [22, 23]. As the “texture” contained in our electronic microscopic images is not of a regularly repeating variety, it is not clear whether these features would help in segmenting the images manually. Moreover, the seven features used in [21] for segmentation when applied on electronic microscopic images lead to large calculations and high complexity. In this paper, we will present a hybrid feature extraction method based on wavelet transform and pixel texture parameters for microscopic image regions. Those used parameters are: pixel intensity, mean, and variance. The proposed technique is described in the incoming section.

## 2. HYBRID TECHNIQUE

Although it may suffer from real time problems, texture- based algorithms are essential in different sections (segmentation, registration, denoising... etc.). As previously described, the commonly used features in almost all texture based segmentation methods are: mean, variance, coefficient of correlation, contrast, and homogeneity [21]. Taking all those features leads to large computations and thus more complexity. In this paper, we present an automatic hybrid method for feature extraction. Based on wavelet transform, pixel intensity, image mean, and variance this proposed technique is applied to images acquired

from an electronic microscope. In this type of images, features are of great importance. Moreover, errors are not accepted as applications are very restricted to scientific and research purposes. This algorithm deals with TEM grayscale images. These images are acquired by electronic microscope which are used for characterize the material nanostructure. The proposed hybrid segmentation method consists of two main steps:

- (1). an automatic threshold generator (ATG)
- (2). High efficient multiple- regions segmentation filter.

### 2.1. AUTOMATIC THRESHOLD GENERATOR (ATG)

The proposed method starts with an automatic filter threshold generation. This ATG aims to get exact threshold values used for electronic microscopic input image segmentation. To perform this task, the following step actions are considered:

- An input image histogram (contrast and homogeneity) is first generated for each input electronic microscopic image data
- Image histogram impulse and false data pixels removal is then applied to reduce errors and increase accuracy of threshold selection
- The input image histogram obtained is fitted to a suitable probability distribution function (PDF) to detect its mean and lobe width
- A histogram fingerprint for each input image is then defined

#### 2.1.1. HISTOGRAM GENERATOR

The input image is preprocessed first to enhance its quality. Histogram shows the intensity distribution within the input image. Image histogram serves in observation of homogeneity or non- homogeneity in different image areas, and thus thresholds decision making within the next step, as shown in Fig 1.

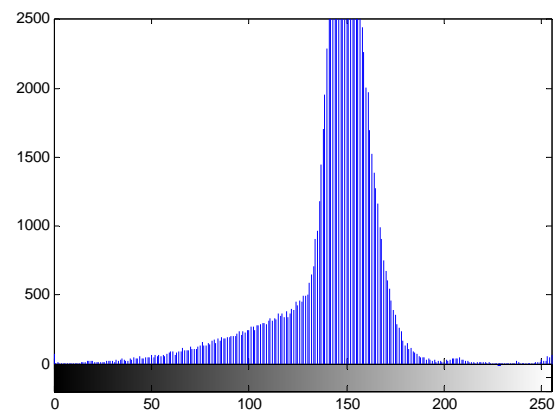


Fig 1: Input image histogram shape.

### 2.1.2. HISTOGRAM IMPULSE DATA FILTRATION

An impulse data filter is used to omit random values within the histogram data to avoid presence of false critical points, which in turns lead to a false threshold selection.

### 2.1.3. HISTOGRAM FITTING METHOD

Observing Fig 1 for the set of input images acquired from the electronic microscope, it could be easily concluded that their histogram data could be fitted to a suitable PDF curve. Two distributions are selected according to the histogram of each input image: the Normal distribution and the Poisson distribution, as shown in “(1)”, “(2)” respectively. The advantage of this fitting method is that both distributions have known mean and widths.

$$I = \frac{1\lambda}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (1)$$

$$I = \frac{\lambda e^{-\lambda}}{k!} \quad (2)$$

The mean and bandwidth of this fitted probability distribution function is recorded. Furthermore, side lobes are to be taken into consideration by critical points determination; critical points are those pixels where maximum or inflection points are found. Gradient methods are applied to the fitted curve to detect these critical points. Results of this gradient method are: centers of side lobes, and inflection points:  $\{(I1, \Delta1), (I2, \Delta2), (I3, \Delta3) \dots\}$ ; Where  $I$  and  $\Delta$  represent the lobe center and lobe width respectively. Since the percentage of infected pixels is usually small enough, infected pixels within the input image are usually away from

the detected lobes (especially the main lobe). Obviously this histogram analysis leads to a good selection of segmentation thresholds or a successful ATG.

### 2.1.4. FEATURE EXTRACTION

An exact method is used to determine threshold values, “Histogram Fingerprint”. What is the meaning of a histogram fingerprint? It is a simple array that contains the exact places and values of histogram critical points (main lobe center- main lobe width- side lobes centers- side lobes widths). Observing these values, we can design the suitable filter accurate parameters: center  $K$  and bandwidth

$N_1$  and  $N_2$ , as in “(3)”. Fig 2 shows the histogram fingerprint details.

$$\begin{aligned} \text{if } K < \frac{n}{2} \quad BW &= k - 1 \\ \text{If } K > \frac{n}{2} \quad BW &= n - k \end{aligned} \quad (3)$$

And  $N_1 = \text{fingerprint}(K - BW)$   
 $N_2 = \text{fingerprint}(K + BW)$

Applying this proposed threshold selection method on a wide data set, it has shown impressive results except when the center of the main lobe ( $K$ ) lies at either one of the ends of the histogram fingerprint. To overcome this problem, we will introduce a binary constant, the ‘deviation factor’ that will be used in both cases for better values of  $N_1$  and  $N_2$ . Equations (4) summarizes the effect of this constant:

$$\begin{aligned} \text{If } K &= 1 \text{ or } 2 \\ BW &= K - 1 \\ N_1 &= \text{fingerprint}(K - BW \pm \text{deviation}) \\ N_2 &= \text{fingerprint}(K + BW) \\ \text{If } K &= n \text{ or } (n - 1) \\ BW &= n - K \\ N_1 &= \text{fingerprint}(K - BW) \\ N_2 &= \text{fingerprint}(K + BW \pm \text{deviation}) \end{aligned} \quad (4)$$

This binary constant moreover add an integer factor of safety called “Deviation Factor” that could be varied

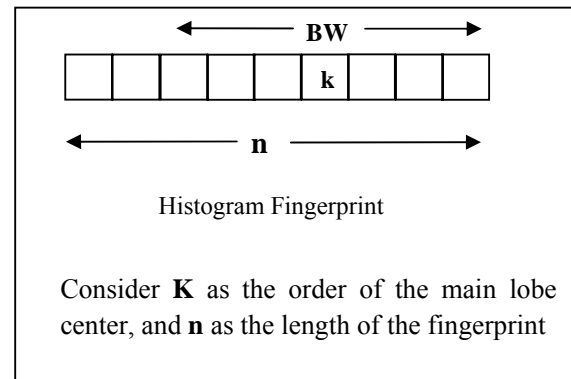


Fig 2 : Histogram Fingerprint.

## 2.2. THE HYBRID ATG- SEGMENTATION PHASE

A region of interest (ROI) is one area or multiple areas of an image of remarkable importance or distinct features from the rest of the image. ROIs are usually filtered for further investigations. Several methods were proposed in this area either for a single ROI [24] or multiple ROIs filtration [25]. We will define existing ROIs by creating one 'binary mask'- a binary image that is the same size as the input image with pixels that define the ROI set to 1 and all other pixels set to 0- for all regions. Infected regions or ROIs are selected according to following procedure:

- Outputs of the histogram fingerprint analyzer (described in the previous section) are first obtained;
- A binary mask is obtained;
- Input image is filtered according to this mask

Fig 3 shows the block diagram of the whole system. It should be noticed that input data set suffers a major segmentation problem; random occurrence of regions. In the incoming section, we will apply this proposed segmentation procedure on a wide data set and observe the results. Moreover, we will verify our proposed methods to segment electronic microscopic images with unexpected number and shapes of regions of interest.

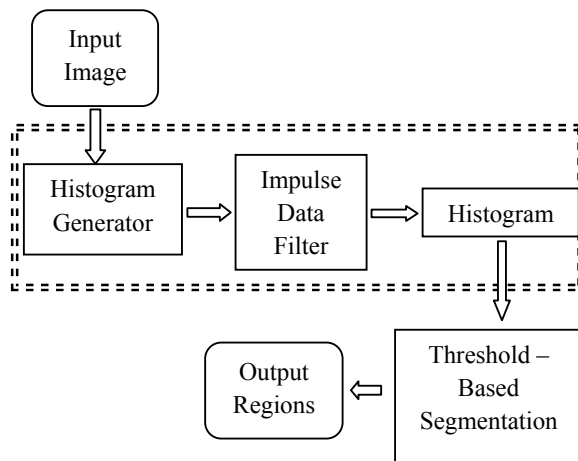


Fig 3: The proposed hybrid ATG- Segmentation method.

## 2.3. PIXEL COUNTING METHOD

Output of the previous section is a binary image, where nano-particles are represented by black pixels and the background is represented by the white areas. Clustering of those black pixels is preformed taking into consideration the connectivity between them. The connectivity will be accomplished using an iterative method. The

method mainly depends on getting certain relation between each pixel and its neighbors. The stopping criterion starts when the clustering results remain unchanged during further iterations. This counting yields to:

- (1). Start with two arrays (A, B) that both have a number of elements corresponding image pixels.
- (2). Fill these arrays with a large integer number, that must be greater than the expected particles number (M)
- (3). Put a variable N as counter for particle Number
- (4). Start N=0
- (5). Calculate every array elements according the next relation

$$B_{i,j} = \text{Min}(A_{i+1,j}, A_{i-1,j}, A_{i,j+1}, A_{i,j-1})$$

- (6). If  $A_{i,j}$  is equal to the larger number M then  $N = N + 1$  and  $A_{i,j} = N$
- (7). If  $A_{i,j}$  is less than M ,  $A_{i,j}$  remains unchanged
- (8). Getting  $r = |A - B|$
- (9). If r equals zero, the solution is accomplished. Else repeat steps 5,6,7,8,9 after putting  $A = B$

## 2.4. PARTICLE DIAMETER

The area of particle is denoted by (S), S is calculated by multiplying the number of pixels contained in the particle by the scale factor. If the particle area S is considered as circle the diameter (D) can be calculated according next relation

$$D = 2\sqrt{\frac{S}{\pi}}$$

Where  $\pi$  is equal 3.14

## 3. RESULTS AND DISCUSSION

### 3.1. THE SAMPLE RESOURCES

The samples of TEM are obtained from obtain from the literatures surveying such as Ref. [8] and Ref. [26] as shown in Fig 4 (a) (b) (c) (d) (e). They represent a variety of image of TEM image with different conditions.

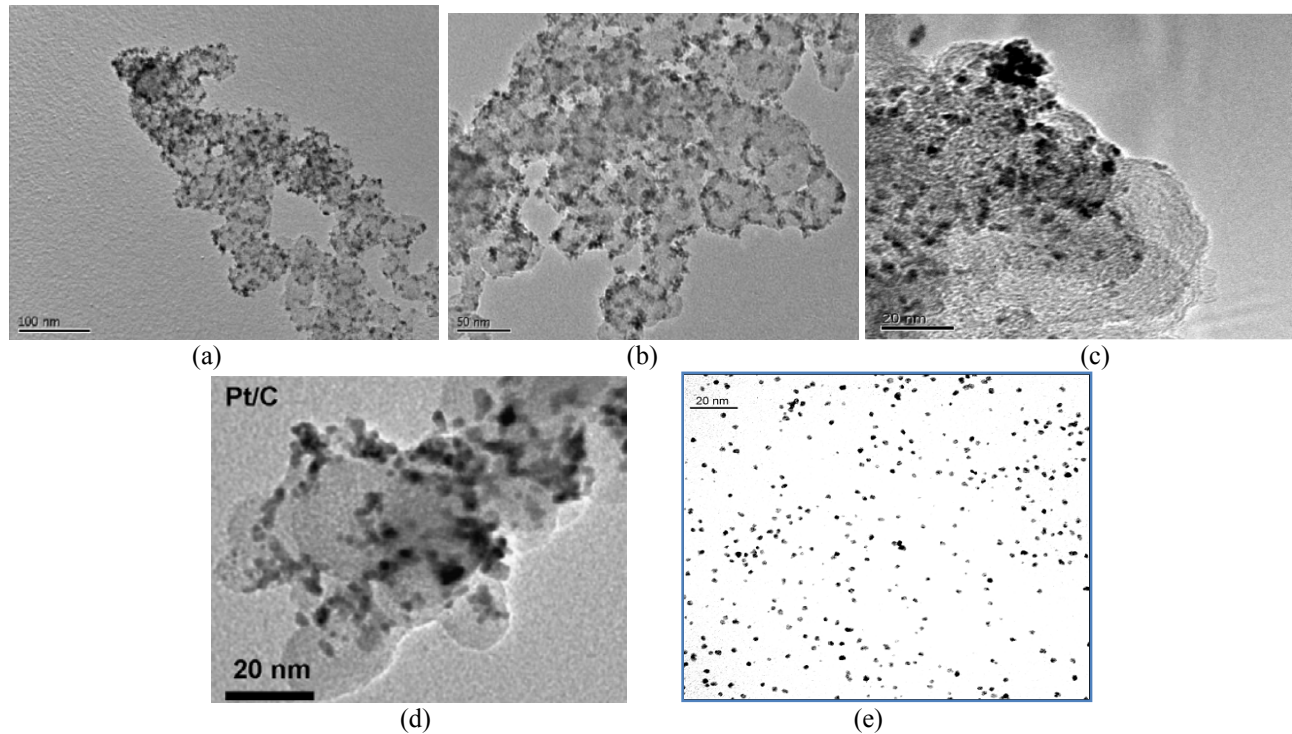


Fig 4 Sample Case studies chosen from [8 and 26].

### 3.2. RESULTS OF THE PROPOSED THRESHOLD- BASED SEGMENTATION METHOD

It is applied to a wide data set. Images acquired from an electronic microscope are all used. Fig 5 shows segmentation method by applying threshold values

acquired when fitting the histogram curve to normal and Poisson PDF respectively. Observing several results, it can be concluded that the histogram curve fitted to normal distribution will give more realistic results. All incoming tests are performed by fitting to normal distribution

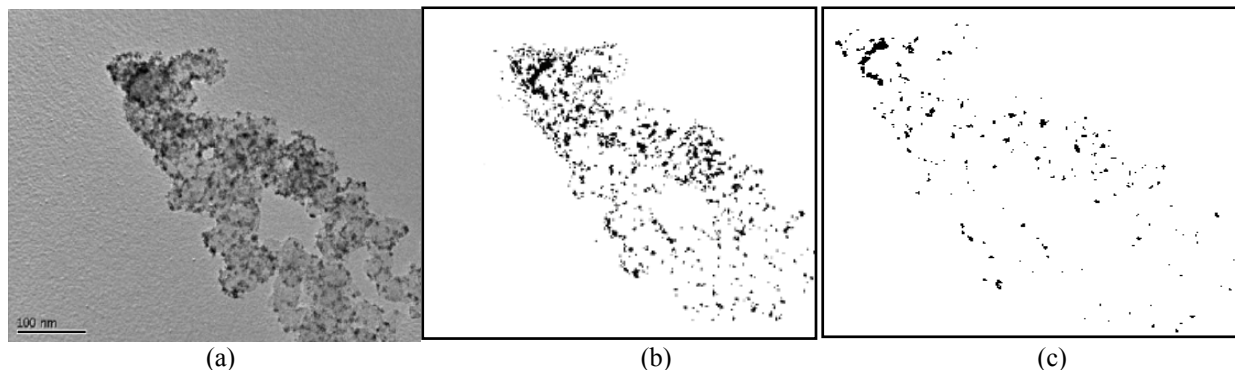


Fig 5 : Segmentation method with threshold values acquired by fitting the histogram curve to normal and Poisson PDF respectively (a) TEM image (b) TEM image processed according to Normal distribution (c) TEM image processed according to Poisson distribution. Filter edges (N1=80 and N2=190) grayscale level.

Fig 6 shows the processed TEM image of Fig 4 (a) [8] using the proposed algorithm and the associated particle size histogram. This case was chosen as an example of TEM images with a partially unclear background and the particles have some variation of intensity; hence, the histogram fingerprint

distribution indicates one main lobe at 148, the filter edges are at (N1=110 and N2=190) and the deviation factor are equal to zero. Fig 7 shows the processed TEM image of Fig 4 (b) [8] processed image using our algorithm and the associated particle size histogram. In this case, the histogram fingerprint

distribution indicates one main lobe at 148, the filter edges are at (N1=128 and N2=244) and the deviation factor are equal to zero, 510 particles could be counted against manually counted of nearly 600 particles. Fig 8 shows the processed TEM image of Fig 4 (c) [8] and the associated particle size histogram. This case was chosen as an example of

TEM images with a partially unclear background and the particles have some variation of intensity; hence, the histogram fingerprint distribution indicates one main lobe at 174, the filter edges are at (N1=174 and N2=220) grayscale level and the deviation factor are equal to zero. 70 particles could be counted against manually counted of 79 particles.

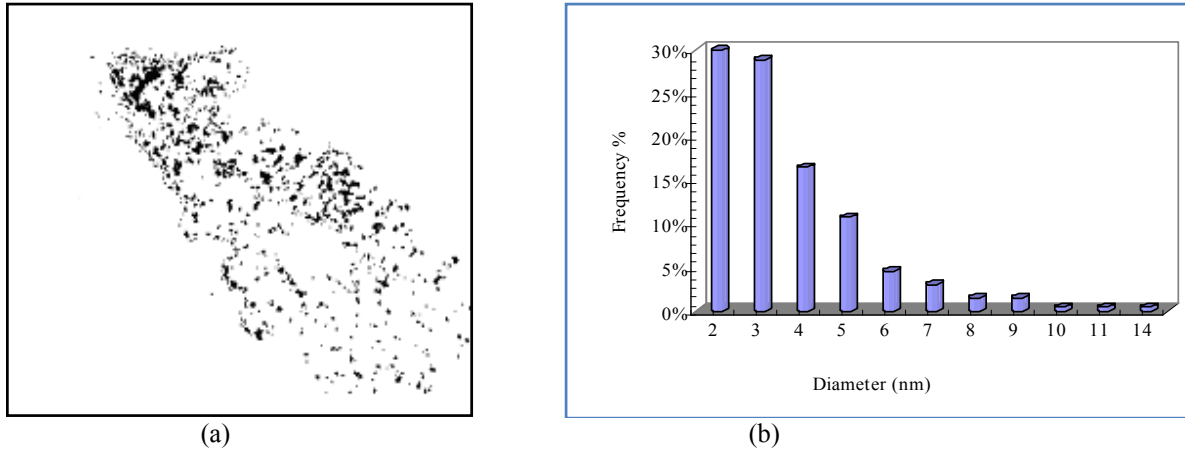


Fig 6: (a) Processed TEM image of Fig 4 (a), (b) associated particle size histogram

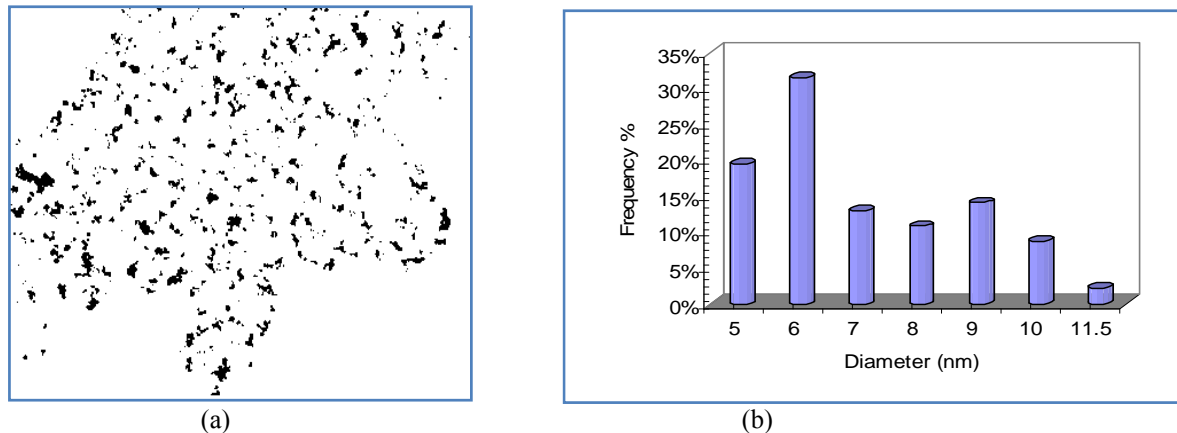


Fig 7: (a) Processed TEM image of Fig 4 (a) (b) associated particles size histogram.

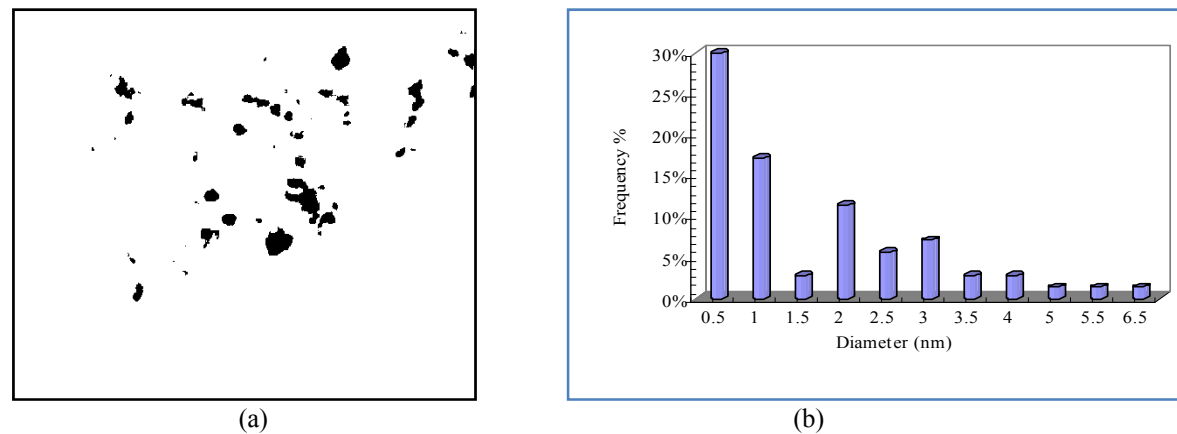
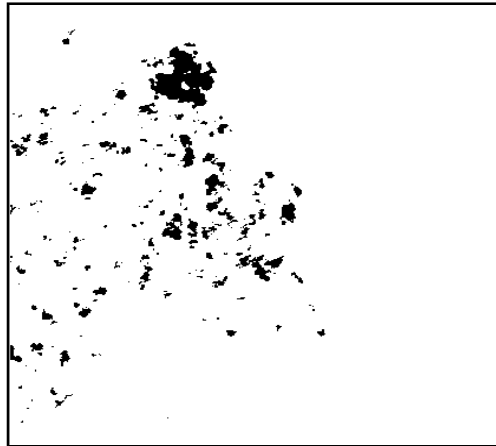


Fig 8 : (a) Processed TEM image of Fig 4 (c), (b) associated particles size histogram.

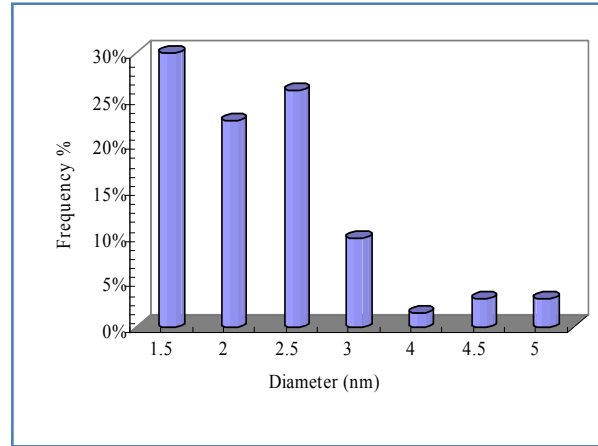


Fig 9 shows the processed TEM image of Fig 4 (d) [8] and the associated particle size histogram. This case was chosen as an example of TEM images with a unclear background and the particles have some variation of intensity; hence, the histogram



(a)

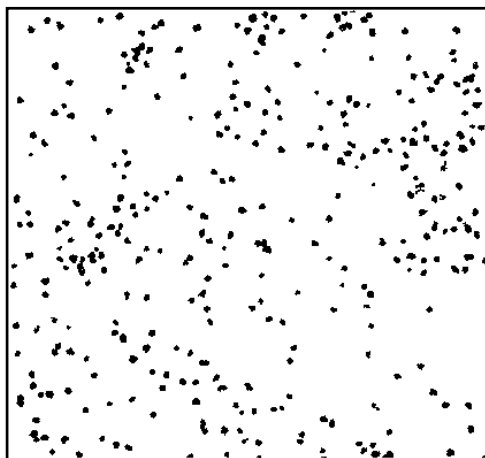
fingerprint distribution indicates one main lobe at 174, the filter edges are at (N1=114 and N2=244) grayscale level and the deviation factor are equal to one. 116 particles could be counted against manually counted of nearly 150 particles.



(b)

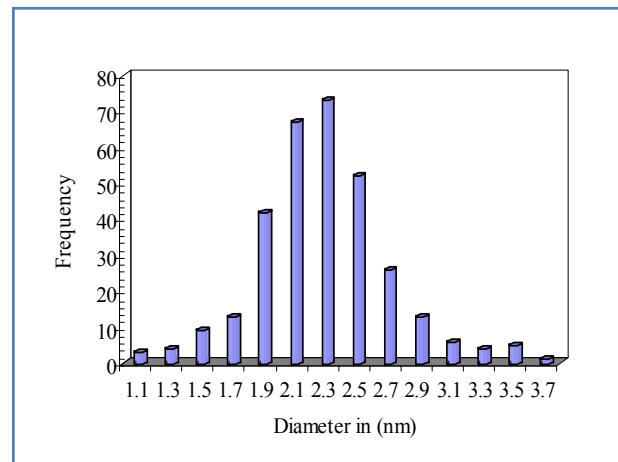
Fig 9: (a) Processed TEM image of Fig 4 (d) using the proposed algorithm (b) Associated particles size histogram.

Fig 10 shows the processed TEM image of Fig 4 (e) [26] and the associated particle size histogram. This case was chosen as an example of TEM images with a clear background but the particles have big variation of intensity; hence, the histogram fingerprint distribution indicates one main lobe at



(a)

174, the filter edges are at (N1=128 and N2=244) grayscale level and the deviation factor are equal to zero. 315 particles could be counted against 369 particles counted in [26]. When the counted manually were found 333 particles.



(b)

Fig 10: (a) Processed TEM image of Fig 4 (e) using the proposed algorithm (b) associated particles size histogram.

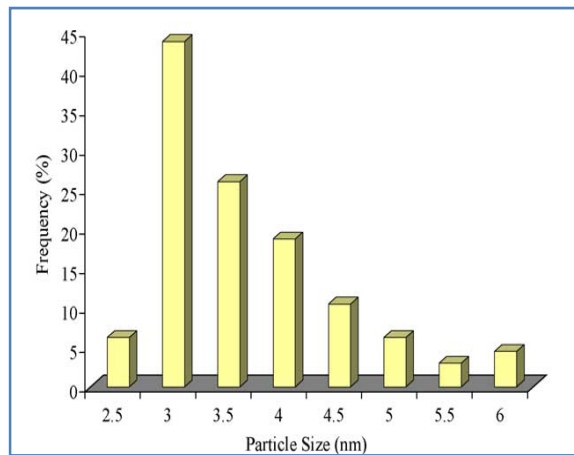
### 3.3. RESULTS COMPARISON

The comparison has been preformed based on the results of Refs. [8, 26]. The case of Fig 4 (c) from Ref. [8] and the case of Fig 4 (e) from [26] are chosen to perform the comparison. Ref [8] has proposed a manual method to count and size the nano

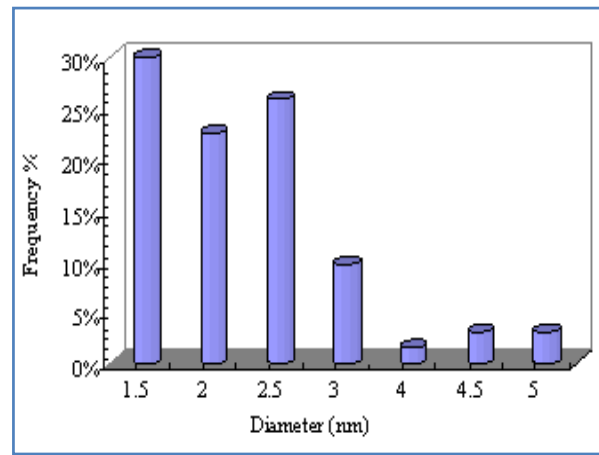
particles where were obtained over 100 particles in randomly chosen areas not over the all image. The Fig 11 shows the particle size histogram of [8] against the particle size histogram using our algorithm. The results which included in Fig 11 (a) are not reasonable because as we see in the figure the

summation of the frequency percentage over the all

registered particle size is nearly equal to 135%.



(a)

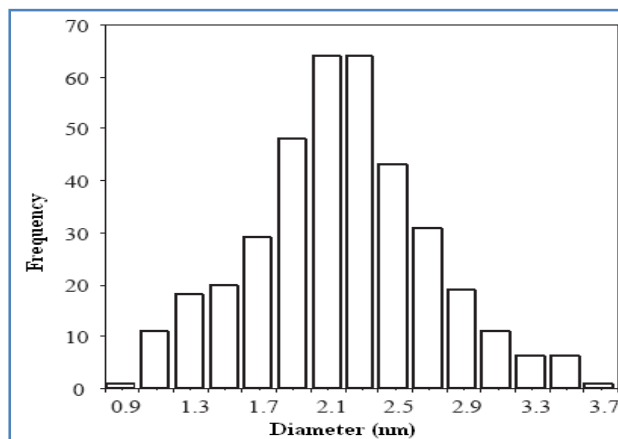


(b)

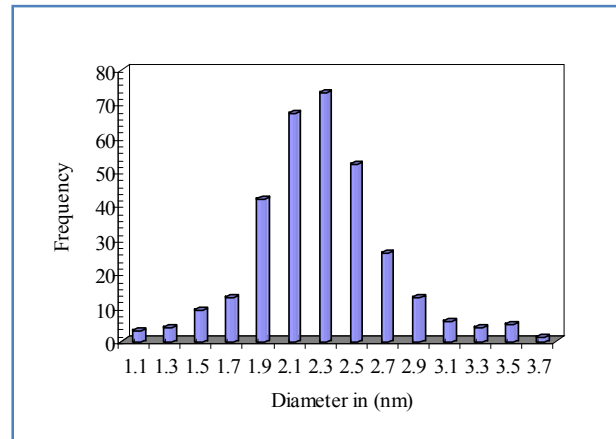
Fig 11: Particle size histogram (a) of Ref. [8] (b) The proposed algorithm.

Ref. [26] has proposed that the contrast and darkness of the separate sections of the TEM image were adjusted in Adobe Photoshop prior to the analyzing of particles by NIH-Image. The Fig 12 shows the particle size histogram of Ref. [26] against the particle size histogram using our algorithm. 315 particles could be counted against 369 particles

counted in [26] and 333 particles manually counted. The results which included in Fig 12 gives an indication that the particles size has almost the same distribution trend. Fig 12 (a) shows that the standard deviation and the mean diameter are 0.437 nm and 2.4 nm respectively while Fig 12 (a) shows 0.54 nm of standard deviation and 2.3 nm of mean diameter.



(a)



(b)

Fig 12: Particle size histogram (a) Ref. [26] (b) The proposed algorithm.

#### 4. CONCLUSION

The presented method for nano-particles size characterization in TEM using the fast hybrid automatic threshold segmentation and counting, shows promising results. Since the algorithm is performed automatically without human interaction, fidelity and better accuracy are both achieved. Moreover, this fast algorithm gives more advantage in execution time. Many cases from [8 and 26] have been tested; comparison between manual, existing,

and proposed techniques shows the impact of our work.

#### REFERENCES

- [1]. Heath B. Henninger, Steve A. Maas, Clayton J. Underwood, Ross T. Whitaker, JeVrey A. Weiss "Spatial distribution and orientation of dermatan sulfate in human medial collateral ligament" Journal of Structural Biology, PP13, 6 November 2006.

- [2]. R. Schaeublin, M.-J. Caturla, M. Wall, T. Felter, M. Fluss, B.D. Wirth, T. Diaz de la Rubia, M. Victoria "Correlating TEM images of damage in irradiated materials to molecular dynamics simulations", *Journal of Nuclear Materials* 307–311, 988–992, 2002.
- [3]. David C. Bell, Yue Wu, Carl J. Barrelet, Silvija Gradecak, Jie Xiang, Brian P. Timko, Charles M. Lieber "Imaging and Analysis of Nanowires", *Microscopy Research And Technique* 64:373–389 (2004).
- [4]. G.W. Chinthaka Silva, Longzhou Ma, Oliver Hemmers, Dennis Lindle "Micro-structural characterization of precipitation-synthesized fluorapatite nano-material by transmission electron microscopy using different sample preparation techniques", *Science Direct, Micron (the international research and review journal for microscopy)*, vol.39 No. 3, 269–274,(2008).
- [5]. Yeng-Ming Lam, Gerhard Goldbeck-Wood, Chris Boothroyd "Mesoscale Simulation and cryo-TEM of Nanoscale DrugDelivery Systems" *Molecular Simulation*, Vol. 30 (4), pp. 239–247 , 15April 2004.
- [6]. Kurt Scheerschmidt and Peter Werner "Analysis of Nanostructures by EM Techniques: Quantum Dots", *EUREM 12*, Brno, Czech Republic, July 9-14, 2000.
- [7]. Kenta Yoshida, Tomoyuki Kawai, Takahiro Nambara, Sakae Tanemura, Koh Saitoh, Nobuo Tanaka "Direct observation of oxygen atoms in rutile titanium dioxide by spherical aberration corrected high-resolution transmission electron microscopy", *Nanotechnology* 17 3944–3950, 2006.
- [8]. Wang Y.C., Chou T.M., Libera M. "Preparation of carbon-supported PtRu nanoparticles for direct methanol fuel cell applications – a comparative study", *Journal of Power Sources* ,Vol. 142, Pp. 43–49 , 2005.
- [9]. Jae H. Kim, Sung M. Choi, Sang H. Nam, Min H. Seo, Sun H. Choi, Won B. Kim, "Influence of Sn content on PtSn/C catalysts for electrooxidation of C1–C3 alcohols: Synthesis, characterization, and electrocatalytic activity", *Applied Catalysis B: Environmental*, Vol. 82, Pp. 89–102, 2008.
- [10]. Atkins, M.S., Mackiewicz, B.T., Fully automatic segmentation of the brain in MRI. *IEEE Trans. Med. Imaging*, Vol. 17, Pp. 98–107. ,1998.
- [11]. Ravinda G. N. Meegama, Jagath C. Rajapakse "Fully Automated Peeling Technique For T1-Weighted, High-Quality MR Head Scans", *International Journal of Image and Graphics (IJIG)*, Volume: 4, Issue: 2, Pp. 141-156,2004.
- [12]. Horst K. Hahn, Heinz-Otto Peitgen . "The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform.", *Proc. MICCAI, LNCS 1935*: Pp. 134-143, Springer, Berlin, 2000.
- [13]. Tina Kapur, W. Eric L. Grimson, William M. Wells, Ron Kikinis2 "Segmentation of Brain Tissue from Magnetic Resonance Images." *Medical Image Analysis* volume 1, number 2, Pp. 109–127, 1996.
- [14]. David W. Shattuck, Stephanie R. Sandor-Leahy, Kirt A. Schaper, David A. Rottenberg, Richard M. Leahy "Magnetic resonance image tissue classification using a partial volume model.", *NeuroImage* 13 (5), Pp. 856–876, 2001.
- [15]. Xu C., Pham D. L., Rettmann M. E., Yu D. N., Prince J. L., "Reconstruction of the Human Cerebral Cortex from Magnetic Resonance Images", *IEEE Transactions on Medical Imaging*, 18(6), pp. 467-480, June, 1999.
- [16]. Zeng, X., Lawrence H. Staib, Robert T. Schultz, James S. Duncan. "Segmentation and Measurement of the Cortex from 3D MR Images Using Coupled Surfaces Propagation". *IEEE Transactions On Medical Imaging*, Vol. 18, No. 10, October 1999.
- [17]. Kim Y., Kim J., Won Y., Yongho In, "Segmentation Of Protein Spots in 2d Gel Electrophoresis Images With Watersheds Using Hierarchical Threshold" , *LNCS - Computer and Information Sciences - ISCIS 2003*, 2869:389-396, 2003.
- [18]. Takahashi, K., Nakazawa M., Watanabe Y., "Dnainsight: An Image Processing System For 2-D Gel Electrophoresis Of Genomic DNA", *Genome Informatics*, Vol. 8, Pp.135-146, 1997.
- [19]. Umer Z. Ijaz , Safee U. Chaudhary, Moon S. Don, Kyung Y. Kim, "Computational Strategies for Protein Quantitation in 2D Electrophoresis Gel Image Processor for Matlab", *Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies*, 129-134, 2007.
- [20]. Christopher S. Hoeflich , Jason J. Corso, "Segmentation of 2D Gel Electrophoresis Spots Using a Markov Random Field", In

- Proceedings of SPIE Conference on Medical Imaging, 2009.
- [21]. A. Solberg and A. Jain, "Texture fusion and feature selection applied to SAR imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 35, no. 2, pp. 475–479, 3 1997.
  - [22]. D. Geman, C. Geman, C. Graffigne, and D. Pong, "Boundary detection by constrained optimization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, pp. 609–628, Jul 1990.
  - [23]. R.M. Haralick, "Statistical and structural approaches to texture," Proc. IEEE, vol. 67, no. 5, pp. 786–804, May 1979.
  - [24]. M. Abdou, M. Tayel, "Automated Biomedical System for Image Segmentation and Transmission," The International Journal of Robotics and Automation, Vol. 23, Issue 1, 2008.
  - [25]. M. Abdou, "An Introduced Blocked-Wavelet Algorithm for Signal and Time Improvements of Images with Multiple Regions," CGIM 2008, 10th IASTED International Conference on Computer Graphics and Imaging, Innsbruck, Austria, pp.34- 39, Feb. 2008.
  - [26]. H. Woehrle, E. Hutchison, S.Ozkar2, G. Finke "Analysis of Nanoparticle Transmission Electron Microscopy Data Using a Public-Domain Image-Processing Program, Image" © Tubitak ,Turk J Chem vol. (30) P.p. 1-13,(2006).

#### AUTHORS PROFILE



**Dr. Mohamed Abd-ElRahman Abdou** is an assistant professor in Communications Engineering. He obtained his PhD in 2006 when he started a tenure track as a researcher at Mubarak City for Sciences & Technology. In 2008 he joined Pharos University in Alexandria as an assistant professor in the EE department. His research interests are in the area of image and signal processing, biomedical engineering, and

knowledge based systems (fuzzy control- Neural Network-cellular automata). He published his first international book in February 2010 in medical image processing. His is a reviewer of several international journals: ELSEVIER, IASTED, WSEAS



**Dr. Bayumy A. B. Youssef** is a researcher in computer graphics department, Information technology institute, Mubarak City for Scientific Research and Technological Applications, Alexandria, Egypt. He received Bs.c. and Ms.c. degrees from Mechanical Engineering department, faculty of Engineering, Cairo University, Egypt. He received his Ph.D degree from the Information Technology department, institute of graduate studies and research, Alexandria University 2004. His research interests include Numerical modeling and Simulation, Visualization of Scientific Data, Mesh generation product, Investigating Conservation and Restoration of Culture Heritage and (2D & 3D) Image Processing (Segmentation, Restoration, & registration)



**Walaa M. Sheta** is an associate professor of Computer graphics in Informatics Research Institute at Mubarak city for Scientific Research (MUCSAT) since 2006. During 2001-2006 he has worked as Assistant professor at MUCSAT. He holds a visiting professor position at University of Louisville in US and University of Salford in UK. He advised approximately 20 master's and doctoral graduates, his research contributions and consulting spans the areas of real-time computer graphics, Human computer Interaction, Distributed Virtual Environment and 3D image processing. He participated and led many national and multinational research funded projects. He received M.Sc. and PhD in Information Technology from University of Alexandria, in 1993 and 2000, respectively. He received B.Sc. from Faculty of Science, University of Alexandria in 1989

# Gaussian Process Model for Uncertain Data Classification

G.V.Suresh<sup>1</sup>

Assoc.Professor, CSE Department  
Universal College of Engineering  
Guntur, India  
vijaysuresh.g@gmail.com

E.V.Reddy<sup>2</sup>

Assoc.Professor, CSE Department  
Universal College of Engineering  
Guntur, India  
ucet\_evr@yahoo.com

Shabbeer Shaik

Assoc. Professor, MCA Department  
Tirmula College of Engineering  
Guntur, India  
skshabbeer@yahoo.com

Usman Ali Shaik

Assoc.Professor, CSE Department  
Universal College of Engineering  
Guntur, India  
usman\_india@yahoo.com

**Abstract**— Data uncertainty is common in real-world applications due to various causes, including imprecise measurement, network latency, out-dated sources and sampling errors. These kinds of uncertainty have to be handled cautiously, or else the mining results could be unreliable or even wrong. We propose that when data mining is performed on uncertain data, data uncertainty has to be considered in order to obtain high quality data mining results. In this paper we study how uncertainty can be incorporated in data mining by using data clustering as a motivating example. We also present a Gaussian process model that can be able to handle data uncertainty in data mining.

**Keywords**- Gaussian process, uncertain data, Gaussian distribution, Data Mining

## I. INTRODUCTION

Data is often associated with uncertainty because of measurement inaccuracy, sampling discrepancy, outdated data sources, or other errors. This is especially true for applications that require interaction with the physical world, such as location-based services [1] and sensor monitoring [3]. For example, in the scenario of moving objects (such as vehicles or people), it is impossible for the database to track the exact locations of all objects at all-time instants. Therefore, the location of each object is associated with uncertainty between updates [4]. These various sources of uncertainty have to be considered in order to produce accurate query and mining results. We note that with uncertainty, data values are no longer atomic. To apply traditional data mining techniques, uncertain data has to be summarized into atomic values. Taking moving-object applications as an example again, the location of an object can be summarized either by its last recorded location or by an expected location. Unfortunately, discrepancy in the summarized recorded value and the actual values could seriously affect the quality of the mining results. In recent years, there is significant research interest in data uncertainty management. Data uncertainty can be categorized into two types, namely existential uncertainty and value uncertainty. In the first type it is uncertain whether the object or data tuple

exists or not. For example, a tuple in a relational database could be associated with a probability value that indicates the confidence of its presence. In value uncertainty, a data item is modelled as a closed region which bounds its possible values, together with a probability density function of its value. This model can be used to quantify the imprecision of location and sensor data in a constantly-evolving environment.

### A. Uncertain data mining

There has been a growing interest in uncertain data mining[1], including clustering[2], [3], [4], [5], classification[6], [7], [8], outlier detection [9], frequent pattern mining [10], [11], streams mining[12] and skyline analysis[13] on uncertain data, etc. An important branch of mining uncertain data is to build classification models on uncertain data. While [6], [7] study the classification of uncertain data using the support vector model, [8] performs classification using decision trees. This paper unprecedentedly explores yet another classification model, Gaussian classifiers, and extends them to handle uncertain data. The key problem in Gaussian process method is the class conditional density estimation. Traditionally the class conditional density is estimated based on data points. For uncertain classification problems, however, we should learn the class conditional density from uncertain data objects represented by probability distributions.

## II. RESEARCH BACKGROUND

### A. General Structure of Gaussian Process Models

The conditional distribution  $p(y|x)$  describes the dependency of an observable  $y$  on a corresponding input  $x \in \mathcal{X}$ . The class of models described in this section assumes that this relation can be decomposed into a systematic and a random component. Further- more, the systematic dependency is given by a latent function  $f: X \rightarrow \mathbb{R}$  such that the sampling distribution, i.e. the likelihood, is of the form

$$p(y|(f(x), \theta)) \quad (2.1)$$



Which describes the random aspects of the data-generating process. In general, we will use to denote additional parameters of the likelihood besides  $f$ . Note that the conditional distribution of the observable  $y$  depends on  $x$  via the value of the latent function  $f$  at  $x$  only. The aim of inference is to identify the systematic component  $f$  from empirical observations and prior beliefs. The data comes in the form of pair-wise observations  $D = \{(y_i, x_i)\} | i = 1, \dots, m\}$  where  $m$  is the number of samples. Let  $X = [x_1, \dots, x_m]^T$  and  $y = [y_1, \dots, y_m]^T$  collect the inputs and responses respectively. In general we assume  $x \in \mathbb{R}^n$  unless stated otherwise. The name Gaussian process model refers to using a Gaussian process (GP) as a prior on  $f$ . The Gaussian process prior is non-parametric in the sense that instead of assuming a particular parametric form of  $f(x, \theta)$  and making inference about  $\theta$ , the approach is to put a prior on function values directly. Each input position  $x \in \mathcal{X}$  has an associated random variable  $f(x)$ . A Gaussian process prior on  $f$  technically means that a priori the joint distribution of a collection of function values  $f$  associated with any collection of  $m'$  inputs  $[x_1, \dots, x_{m'}]^T$  is multivariate normal (Gaussian)

$$p(f | X, \psi) = N(f | m, K) \quad (2.2)$$

with mean  $m$  and covariance matrix  $K$ . A Gaussian process is specified by a mean function  $m(x)$  and a covariance function  $k(x, x', \psi)$  such that  $K_{ij} = k(x_i, x_j, \psi)$  and  $m = [m(x_1), \dots, m(x_m)]^T$ . By choosing a particular form of covariance function we may introduce hyper-parameters  $\psi$  to the Gaussian process prior. Depending on the actual form of the covariance function  $k(x, x', \psi)$  the hyper-parameters  $\psi$  can control various aspects of the Gaussian process. The sampling distribution (2.1) of  $y$  depends on  $f$  only through  $f(x)$ . As an effect the likelihood of  $f$  given  $D$  factorizes

$$p(y | f, \theta) = \prod_{i=1}^m p(y_i | f(x_i), \theta) = p(y | f, \theta) \quad (2.3)$$

and depends on  $f$  only through its value at the observed inputs  $f$ . According to the model, conditioning the likelihood on  $f$  is equivalent to conditioning on the full function  $f$ . This is of central importance since it allows us to make inference over finite dimensional quantities instead of handling the whole function  $f$ . The posterior distribution of the function values  $f$  is computed according to Bayes' rule

$$p(f | D, \theta, \psi) = \frac{p(y | f, \theta) p(f | X, \psi)}{p(D | \theta, \psi)} = \frac{N(f | m, K)}{p(D | \theta, \psi)} \prod_{i=1}^m p(y_i | f_i, \theta) \quad (2.4)$$

Where  $f_i = f(x_i)$ . The posterior distribution of  $f$  can be used to compute the posterior predictive distribution of  $f(x_*)$  for any input  $x_*$  where in the following the asterisk is used to mark test examples. If several test cases are given,  $X_*$  collects the test inputs and  $f_*$  denotes the corresponding vector of latent function values. The predictive distribution of  $f_*$  is obtained by integration over the posterior uncertainty

$$p(f_* | D, X_*, \theta, \psi) = \int p(f_* | f, X, X_*, \psi) p(f | D, \theta, \psi) df \quad (2.5)$$

Where the first term of the right hand side describes the dependency of  $f_*$  on  $f$  induced by the GP prior. The joint prior distribution of  $f$  and  $f_*$  due to the GP prior is multivariate normal

$$p(f_* | f, X, X_*, \psi) = N \left( \begin{bmatrix} f \\ f_* \end{bmatrix} \middle| \begin{bmatrix} m \\ m_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \quad (2.6)$$

Where the covariance matrix is partitioned such that  $K_{**}$  is the prior covariance matrix of the  $f_*$  and  $K_*$  contains the covariances between  $f$  and  $f_*$ . The conditional distribution of  $f_* | f$  can be obtained from the joint distribution (2.6) using relation to give

$$p(f_* | f, X, X_*, \psi) = N(f_* | m_* + K_*^T K^{-1} (f - m), K_{**} - K_*^T K^{-1} K_*) \quad (2.7)$$

which is again multivariate normal. The simplest possible model assumes that the function can be observed directly  $y = f(x)$  so that the posterior on  $f$  becomes  $p(f | D) = \delta(f - y)$ , describing that no posterior uncertainty about  $f$  remains. From this posterior the predictive distribution of  $f_*$  can be obtained according to eq. (2.5) which corresponds to simply replacing  $f$  by  $y$  in eq. (2.7) Figure 1 shows the posterior Gaussian process which is obtained by conditioning the prior on the five observations depicted as points. The predictive uncertainty is zero at the locations where the function value has been observed. Between the observations the uncertainty about the function value grows and the sampled functions represent valid hypothesis about  $f$  under the posterior process.

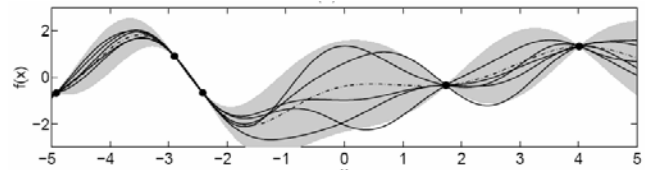


Figure 1 show the posterior Gaussian process on the five observations

### III. OVERVIEW OF THE PROBLEM

We assume the following statistical model

$$t = f(x) + \varepsilon_t \quad (3.1)$$

where  $x$  is a  $D$ -dimensional input and  $\varepsilon_t$  the output, additive, Gaussian uncertain data such that  $\varepsilon_t \sim N(0, v_t)$ , where  $v_t$  is the unknown data variance. Such a model implies that

$$E[t | x] = f(x) \quad (3.2)$$

Now, let  $x = u + \varepsilon_x$  or  $x \sim N(u, v_x I)$  where  $I$  is the  $D \times D$  identity matrix and  $v_x$  is the input data variance. In this case, the expectation of  $t$  given the characteristics of  $x$  is obtained by integrating over the input distribution

$$E[t | u, v_x] = \int f(x) p(x) dx \quad (3.3)$$

This integral cannot be solved analytically without approximations for many forms of  $f(x)$

#### A. Analytical approximation using the Delta-Method

The function  $f$  of the random argument  $x$  can always be approximated by a second order Taylor expansion around the mean  $u$  of  $x$ :

$$f(x) = f(u) + (x-u)^T f'(u) + \frac{1}{2}(x-u)^T f''(u)(x-u) + O(\|x-u\|^3) \quad (3.4)$$

where  $f'(u) = \frac{\partial f(x)}{\partial x}$ , and  $f''(u) = \frac{\partial^2 f(x)}{\partial x \partial x^T}$ , evaluated at  $x=u$ . Within this approximation, we can now solve the (3.3) integral. We have

$$E[t|u, v_x] = \int \left[ f(u) + (x-u)^T f'(u) + \frac{1}{2}(x-u)^T f''(u)(x-u) \right] p(x) dx$$

$$= f(u) + \frac{1}{2} \text{Tr}[f''(u) v_x I] = f(u) + \frac{v_x}{2} \text{Tr}[f''(u)] \quad (3.5)$$

Where  $\text{Tr}$  denotes the trace. Thus, the new generative model for our data

$$\begin{cases} t = g(u, v_x) + \varepsilon_t \\ g(u, v_x) = f(u) + \frac{v_x}{2} \text{Tr}[f''(u)] \end{cases} \quad (3.6)$$

## IV. RELATED WORKS

#### A. Defining a new Gaussian Process

In the case of uncertain or random inputs, the new input/output relationship is given by (3.6), where the former function  $f$ , in the noise-free case, has been replaced by

$$g(u, v_x) = f(u) + \frac{v_x}{2} \text{Tr}[f''(u)].$$

If we put a Gaussian prior on  $f(u)$ , we can derive the corresponding prior on its second derivative and then define the prior on the space of admissible functions  $g(u, v_x)$  which is viewed as the sum of the two

correlated random functions,  $f(u)$  and  $\frac{v_x}{2} \text{Tr}[f''(u)]$ . We use

results from the theory of random functions [3.5]. Let us recall that if  $X(r)$  and  $Y(r)$  are two random functions of the same argument  $r$ , with expected values  $m_x(r)$  and  $m_y(r)$  and covariance functions  $C_x(r, r')$  and  $C_y(r, r')$  respectively, then the mean and covariance function of  $Z(r) = X(r) + Y(r)$  are given by

$$m_z(r) = m_x(r) + m_y(r) \quad (4.1)$$

$$C_z(r, r') = C_x(r, r') + C_y(r, r') + C_{xy}(r, r') + C_{yx}(r, r') \quad (4.2)$$

in the case  $X(r)$  and  $Y(r)$  are correlated  $C_{xy}(r, r')$  and  $C_{yx}(r, r')$  are the cross-covariance functions. We can now apply this to our function  $g(\cdot)$ . Let us first derive the mean and covariance function of  $g(u, v_x)$  in the one-dimensional

case and then extend these expressions to  $D$  dimensions. Given that  $f(u)$  has zero-mean and covariance function

$$C(u_i, u_j) \text{ as given by } C(u_i, u_j) = v \exp\left(-\frac{1}{2} \sum_{d=1} w_d (u_i^d - u_j^d)^2\right)$$

its second derivative  $f''(u)$  has zero-mean and covariance function  $\partial^4 C(u_i, u_j) / \partial u_i^2 \partial u_j^2$ . It is then straightforward that

$$\frac{v_x}{2} f''(u) \text{ has zero-mean and covariance function}$$

$$\frac{v_x^2}{4} \partial^4 C(u_i, u_j) / \partial u_i^2 \partial u_j^2 \text{ also, the cross-covariance functions}$$

$$\text{between } f(u) \text{ and } \frac{v_x}{2} f''(u) \text{ is given by } \frac{v_x}{2} \partial^2 C(u_i, u_j) / \partial u_i^2.$$

Therefore, using the fact we have  $\frac{\partial^2 C(u_i, u_j)}{\partial u_i^2} = \frac{\partial^2 C(u_i, u_j)}{\partial u_j^2}$ , in one

dimension,  $g(u, v_x) = f(u) + \frac{v_x}{2} f''(u)$  has zero-mean and covariance function

$$\text{cov}[g(u_i, v_x), g(u_j, v_x)] = C(u_i, u_j) + \frac{v_x^2}{4} \frac{\partial^4 C(u_i, u_j)}{\partial u_i^2 \partial u_j^2} + v_x \frac{\partial^2 C(u_i, u_j)}{\partial u_i^2} \quad (4.3)$$

In the case of  $D$ -dimensional inputs, we have

$$\text{cov}[g(u_i, v_x), g(u_j, v_x)] = C(u_i, u_j) + \frac{v_x^2}{4} \text{Tr} \left[ \frac{\partial^2}{\partial u_i \partial u_i^T} \left[ \frac{\partial^2 C(u_i, u_j)}{\partial u_i \partial u_j^T} \right] \right] + v_x \text{Tr} \left[ \frac{\partial^2 C(u_i, u_j)}{\partial u_i \partial u_i^T} \right] \quad (4.4)$$

Where  $\frac{\partial^2}{\partial u_i \partial u_i^T} \left[ \frac{\partial^2 C(u_i, u_j)}{\partial u_i \partial u_j^T} \right]$  is a  $D \times D$  each entry of which

$$\text{being a } D \times D \text{ the block } (r, s) \text{ contains } \frac{\partial^2}{\partial u_i^r \partial u_i^s} \left[ \frac{\partial^2 C(u_i, u_j)}{\partial u_i \partial u_j^T} \right]$$

So we see that the first term of the *corrected* covariance function corresponds to the noise-free case plus two correction terms weighted by the input noise variance, which might be either learnt or assumed to be known *a priori*.

#### B. Inference and prediction

Within this approximation, the likelihood of the data  $\{t_1, \dots, t_N\}$  is readily obtained. We have

$$t|U \sim N(0, Q) \text{ with } Q_{ij} = \sum_{ij} + v t \delta_{ij} \quad (4.5)$$

Where  $t$  is the  $N \times 1$  vector of observed targets.  $U$  the  $N \times D$  matrix of input means,  $\sum_{ij}$  is given by (4.4) and

$\delta_{ij} = 1$  when  $i = j, 0$  otherwise. The parameters

$\Theta = [w_1, \dots, w_D, v, v_x, v_t]$  can then be learnt either in a Maximum Likelihood framework or in a Bayesian way, by assigning priors and computing their posterior distribution.

When using the *usual* GP, the predictive distribution of a model output corresponding to a new input  $u_*$ ,  $p(f(u_*) | \Theta, \{u, t\}, u_*)$  is Gaussian with mean and variance respectively given by

$$\begin{cases} \mu = \mathbf{k}^T \mathbf{Q}^{-1} \mathbf{t} \\ \sigma^2 = k - \mathbf{k}^T \mathbf{Q}^{-1} \mathbf{k} \end{cases} \quad (4.6)$$

where  $\mathbf{k}$  is the vector of covariances between the test and the training inputs and  $\sim$ the covariance

between the test input and itself. We have  $Q_{ij} = \sum_{ij} + v_i \delta_{ij}$  and

$$\sum_{ij} = C(\mathbf{u}_i, \mathbf{u}_j), k = C(\mathbf{u}_*, \mathbf{u}_*) \quad (4.7)$$

for  $i, j = 1, \dots, N$  and with  $C(.,.)$ .

With our new model, the prediction at a new (one-dimensional) noise-free input leads to a predictive mean and variance, again computed using (4.6) but with  $Q_{ij} = \sum_{ij} + v_i \delta_{ij}$  with  $\sum_{ij}$  computed as (4.3), and

$$k_i = C(u_*, u_i) + \frac{v_x}{2} \frac{\partial^2 C(u_*, u_i)}{\partial u_i^2} \quad (4.8)$$

$$k = C(u_*, u_*)$$

thus taking account of the randomness in the training inputs.

With this new model, the prediction at a random input is straightforward, simply by using the *corrected* covariance function to compute the covariances involving the test input. Assuming  $x_* \sim N(u_*, v_x)$  we have

$$k_i = C(u_i, u_*) + \frac{v_x^2}{4} \frac{\partial^4 C(u_i, u_*)}{\partial u_i^2 \partial u_*^2} + v_x \frac{\partial^2 C(u_i, u_*)}{\partial u_i^2} \quad (4.9)$$

$$k = C(u_*, u_*) + \frac{v_x^2}{4} \frac{\partial^4 C(u_*, u_*)}{\partial u_*^2 \partial u_*^2} + v_x \frac{\partial^2 C(u_*, u_*)}{\partial u_*^2}$$

## V. EXPERIMENTS

We have implemented the this approach using Matlab 6.5, on 3 real data sets taken from the UCI Machine Learning Repository i.e. Glass data, Iris, Wine data sets. We compare the classification performance of this model on this UCI datasets

TABLE I. STATISTICS OF THE DATASETS ARE LISTED AS FOLLOWS

	Glass	Iris	Wine
Number of Data	214	150	178
Number of features	10	4	13
Number of Classes	6	3	3

We specify a Gaussian process model as follows: a constant mean function, with initial parameter set to 0, a squared exponential with covariance function This covariance function has one characteristic length-scale parameter for each dimension of the input space, and a signal magnitude parameter, for a total of 3 parameters. We train the hyper-parameters using to minimize the negative log marginal likelihood. We allow for 40 function evaluations, and specify that inference should be done with the Expectation

Propagation (EP) inference method and pass the usual parameters

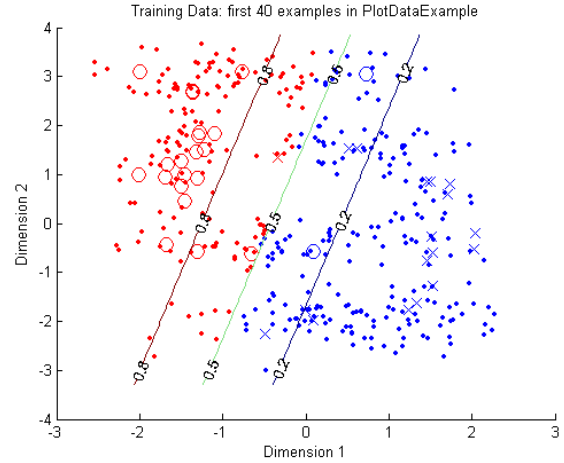


Figure 3: Data plot for the given Data sets

We also Plot the prediction function with the Original data for Glass data

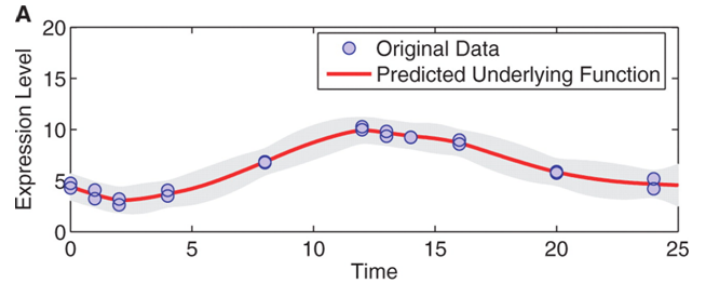


Figure 4: Plot for original data and the Predicted Data

We also plot Log-log plot of the log likelihood of the data against the length scales. The log likelihood is shown as a solid line. The log likelihood is made up of a data fit term (the quadratic form) shown by a dashed line and a complexity term (the log determinant) shown by a dotted line. The data fit is larger for short length scales; the complexity is larger for long length scales. The combination leads to a maximum around the true length scale value of 1.

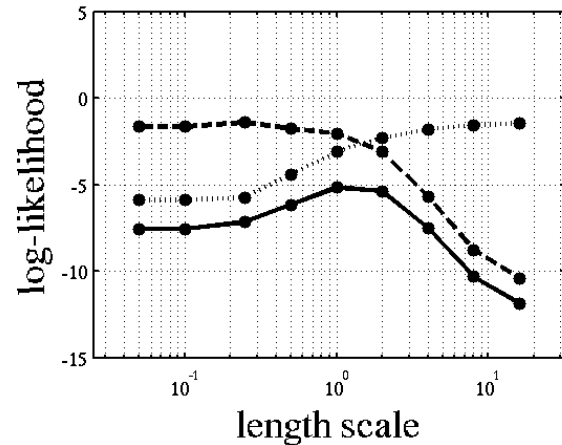


Figure 5: Plot for Length Scale and Log-Likelihood

Therefore, the classification and prediction process is more sophisticated and comprehensive and has the potential to achieve higher accuracy

## VI. CONCLUSION

In this paper, we propose a Gaussian Process model for classifying and predicting uncertain data. The new process is based on an approximation of the random function around the input mean. This process also highlights the correlation between all the parameters, indicating the nature of the likelihood function, and the potential problems for maximum likelihood optimization. We employ the probability distribution which represent the uncertain data attribute, and redesign the Gaussian Process so that they can directly work on uncertain data distributions. We plan to explore more classification approaches for various uncertainty models and find more efficient training algorithms in the future

## REFERENCES

- [1] Aggarwal, C.C.: A Survey of Uncertain Data Algorithms and Applications. IEEE Transactions on Knowledge and Data Engineering 21(5) (2009)
- [2] Cormode, G., McGregor, A.: Approximation algorithms for clustering uncertain data. In: Principle of Data base System, PODS (2008)
- [3] Aggarwal, C.C., Yu, P.: A framework for clustering uncertain data streams. In: IEEE International Conference on Data Engineering, ICDE (2008).
- [4] Singh, S., Mayfield, C., Prabhakar, S., Shah, R., Hambrusch, S.: Indexing categorical data with uncertainty. In: IEEE International Conference on Data Engineering (ICDE), pp. 616–625 (2007)
- [5] Kriegel, H., Pfeifle, M.: Density-based clustering of uncertain data. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 672–677 (2005).
- [6] Bi, J., Zhang, T.: Support Vector Machines with Input Data Uncertainty. In: Proceeding of Advances in Neural Information Processing Systems (2004)
- [7] Aggarwal, R., Imielinski, T., Swami, A.N.: Database mining: A performance perspective. IEEE Transactions on Knowledge & Data Engineering (1993)
- [8] Aggarwal, C.C.: Managing and Mining Uncertain Data. Springer, Heidelberg (2009)
- [9] C. C. Aggarwal and P. S. Yu, "Outlier detection with uncertain data," in *SDM*. SIAM, 2008, pp. 483–493
- [10] Hamdan, H. and Govaert, G. "Mixture Model Clustering of Uncertain Data," *IEEE International Conference on Fuzzy Systems* (2005) 879-884
- [11] Girard, A. & Rasmussen, C. E. & Murray-Smith, R. (2002) Gaussian Process Priors With Uncertain Inputs: Multiple-Step Ahead Prediction. Technical Report, TR-2002-119, Dept. of Computing Science, University of Glasgow.
- [12] Quinonero Candela, J. & Girard, A. (2002) Prediction at an Uncertain Input for Gaussian Processes and Relevance Vector Machines - Application to Multiple-Step Ahead Time-Series Forecasting. Technical Report, IMM, Danish Technical University.
- [13] Rasmussen, C. E. (1996) Evaluation of Gaussian Processes and other Methods for Non-Linear Regression PhD thesis, Dept. of Computer Science, University of Toronto.
- [14] Williams, C. K. I. & Rasmussen, C. E. (1996) Gaussian Processes for Regression Advances in Neural Information Processing Systems 8 MIT Press
- [15] Williams, C. K. I. (2002) Gaussian Processes To appear in The handbook of Brain Theory and Neural Networks, Second edition MIT Press.
- [16] Corinna Cortes and Vladimir N. Vapnik. Support-Vector Networks. Machine Learning, 20(3):273–297, 1995.
- [17] Vladimir N. Vapnik. Statistical Learning Theory. John Wiley & Sons, NY, 1998
- [18] Joaquin Quiñero-Candela and Agathe Girard, "Prediction at an uncertain input for Gaussian processes and relevance vector machines - application to multiple-step ahead time-series forecasting," Tech. Rep., IMM, DTU, 2002, [http://isp.imm.dtu.dk/staff/jqc/prop\\_uncert.ps.gz](http://isp.imm.dtu.dk/staff/jqc/prop_uncert.ps.gz).

## AUTHORS PROFILE

**G.V.Suresh<sup>1</sup>** received his B.Tech and M.Tech from JNT University Hyderabad, India in 2004 and 2008 respectively and currently working toward PhD degree from JNT University. His research interests are in Data Mining Neural Networks, and Networks. He is a life member for Computer Society of India (CSI).

**Shabbeer Shaik** received his MCA and M.Tech from Andhra University India in 1981 and 1985 respectively. His research interests are in Data Mining Neural Networks, and Networks. He is a life member for Computer Society of India (CSI).

**E.V.Reddy<sup>2</sup>** received his M.Sc from Nagarjuna University and he done his M.Tech from Satyabhama University India in 2001 and 2003 respectively. His research interests are in Data Mining Neural Networks, and Image processing. He is a life member for Computer Society of India (CSI).

**Usman Ali Sheik** received his B.Tech and M.Tech from JNT University India in 2003 and 2010 respectively. His research interests are in Data Mining Neural Networks, and Networks. He is a life member for Computer Society of India (CSI).

# Pair Wise Sorting: A New Way of Sorting

Md. Jahangir Alam<sup>1, a</sup>, Muhammad Monsur Uddin<sup>2, b</sup>, Mohammad Shabbir Hasan<sup>3, c</sup>, Abdullah Al Mahmood<sup>4, c</sup>  
E-mail: <sup>1</sup>palash\_14@yahoo.com, <sup>2</sup>dolon\_it@yahoo.com, <sup>3</sup>shabbir\_cse03@yahoo.com, <sup>4</sup>sajibcseofkuet@gmail.com

a. Dhaka International University, Dhaka, Bangladesh

b. Institute of Science, Trade and Technology, Dhaka, Bangladesh

c. Panacea Research Lab, Dhaka, Bangladesh

**Abstract**— This paper presents a technique for sorting numerical data in an efficient way. The numbers of comparisons i.e. the running time of this technique is dependent on distribution or diversity of the value of data items as like as other efficient algorithms. When the total number of data is even, this method groups that data into a collection of pairs and therefore establishes the sorting constraints on each of the pairs. The control is traversed through the list of elements by changing the position of each pair which is the major principle of this technique. On the other hand, when the total number of elements is odd, this method sorts all elements except the last one in the same was as mentioned earlier and the last element is sorted using the general Insertion Sort. This algorithm is therefore a hybrid sorting method that sorts elementary numeric data in a faster and efficient manner.

**Keywords**- Sorting, Pair Wise Sorting, Sorting Techniques.

## I. INTRODUCTION

Sorting is a computational building block of fundamental importance and is one of the most widely studied algorithmic problems [1, 2]. The importance of sorting has also lead to the design of efficient sorting algorithms for a variety of fields like: parallel architectures [3], database systems [4], computer graphics and geographic information systems [5, 6], parallel programming patterns [7, 8] and so on. Many algorithms rely on the availability of efficient sorting routines as a basis for their own efficiency, and many other algorithms can be conveniently phrased in terms of sorting [9]. It is therefore important to provide efficient sorting routines on practically any programming platform.

In this paper, we present a new sorting technique which is based on pair wise comparison. It works separately in two different situations: when total number of elements to be sorted is odd and when total number of elements to be sorted is even. This sorting technique has no major efficiencies over other already existing techniques. But its symmetric structure makes it simpler than other techniques.

This paper is organized as follows. Section 2 describes the proposed technique in details, and the paper continues in Section 3 by calculating complexity of the proposed sorting technique. Section 4 compares the performance of the proposed technique with other already existing sorting techniques. We finish by drawing some broad and necessarily speculative and personal conclusions and future goal in Section 5.

## II. PROPOSED TECHNIQUE- PAIR WISE SORTING

The proposed sorting technique works in two different strategies depending on whether the number of elements to be sorted is odd or even. When the total number of data is even, this method groups that data into a collection of pairs and therefore establishes the sorting constraints on each of the pairs. The control is traversed through the list of elements by changing the position of each pair which is the major principle of this technique. On the other hand, when the total number of elements is odd, this method sorts all elements except the last one in the same was as mentioned earlier and the last element is sorted using the general Insertion Sort. This section describes the proposed sorting technique in details.

### A. Working Procedure:

Let an array contains  $n$  elements where  $n$  is even. To sort these elements the proposed technique uses total  $\frac{n}{2}$  phases where each phase contains two sub phases. The operations performed by two sub phases are different while the functions of all phases are identical. In the first sub phase the algorithm divides the  $n$  elements from position 1 to  $n$  into a total of  $\frac{n}{2}$  pairs. The control moves from first pair to the last pair and checks whether the first element of the pair is larger than the second one and if yes, then these elements are interchanged. In the second sub phase,  $(n-2)$  elements from position 2 to  $(n-1)$  are divided unto a total of  $\frac{n}{2} - 1$  pairs and the similar checking and swapping occur as mentioned earlier.

An additional check is maintained at the end of each phase to detect whether any swapping has occurred or not. If no interchange has been occurred, then the array is declared to be sorted and there is no need to continue for further phases. Otherwise, the phases are continued in the similar fashion till the final phase. Introduction of this checking may increase complexity as it requires initialization, changing and testing during the phases. But it is very effective for the cases where large numbers of elements are to be sorted and elements are uniformly distributed.

This procedure indicates that the number of required phases is constant for a data set and it is half of the total number of



elements. This algorithm is therefore very suitable for sorting large number of elements.

### B. Formulating Algorithm

The proposed technique combines three functions:

1. PAIRWISE (A, N)
2. PROCESS (A, N)
3. INSERT (A, N)

Where A is the array and N is the number of elements to be sorted. These functions are illustrated below:

#### Algorithm: PAIRWISE (A, N)

1. [Initialize] Set oddEvenChk: = N mod 2  
[Detects whether the array contains even number of elements or odd number of elements]
2. [Compare] if oddEvenChk = 0 then  
    Call PROCESS (A, N)  
Else  
    a. Call PROCESS (A, N-1) [Sorts the first N-1 elements]  
    b. Call INSERT (A, N) [Sorts the last N<sup>th</sup> elements]
3. Exit.

#### Algorithm: PROCESS (A, N)

//This function sorts array A with N elements where N is even.

1. [Initialize control variable] Set PTR: = 1
2. [Start of external loop] Repeat Steps from 3 to 8  
While (PTR ≤ N/2)
3. [Initialize variable for first inner loop]  
Set START: =1, END: = N, FLAG: = 0
4. [Start of first inner loop]  
Repeat While (START < END)  
    a. if A[START] < A [START+1] then:  
        swap (A[START], A [START+1])  
        and  
        Set FLAG: = FLAG+1.  
    b. Set START: = START+2  
[End of first inner loop]
5. [Initialize variable for second inner loop]  
Set START: =2, END: = N-1
6. [Start of second inner loop]  
Repeat While (START < END)  
    a. if A[START] < A [START+1] then:  
        swap (A[START], A [START+1])  
        and  
        Set FLAG: = FLAG+1.  
    b. Set START: = START+2  
[End of second inner loop]
7. [Detect whether A is sorted or not]  
If FLAG = 0 then go to Step 9
8. [Increment counter for the external loop]  
Set PTR: = PTR+1  
[End of external loop]
9. Exit

#### Algorithm: PROCESS (A, N)

A is an array with N elements, where N-1 elements are sorted in increasing order. This function finds the actual position of the N<sup>th</sup> element and inserts it into the array by the general insertion sort technique. As the technique is well known, hence it is not mentioned here.

### C. Example

Figure 1 demonstrates an example where an array of eight (8) elements is sorted using the pair wise sorting technique.

Here,

Total number of elements, N = 8

$$\text{Total number of phases} = \frac{N}{2} = \frac{8}{2} = 4$$

Number of comparisons in all passes are same and that is N-1 = 8-1 = 7

In pair wise sorting technique, a flag can be used to cut down the passes, but no flag is used in this example.

$$\text{Total number of pairs in the first sub phase} = \frac{N}{2} = \frac{8}{2} = 4$$

$$\text{Total number of pairs in the second sub phase} = \frac{N}{2} - 1 = \frac{8}{2} - 1 = 4 - 1 = 3$$

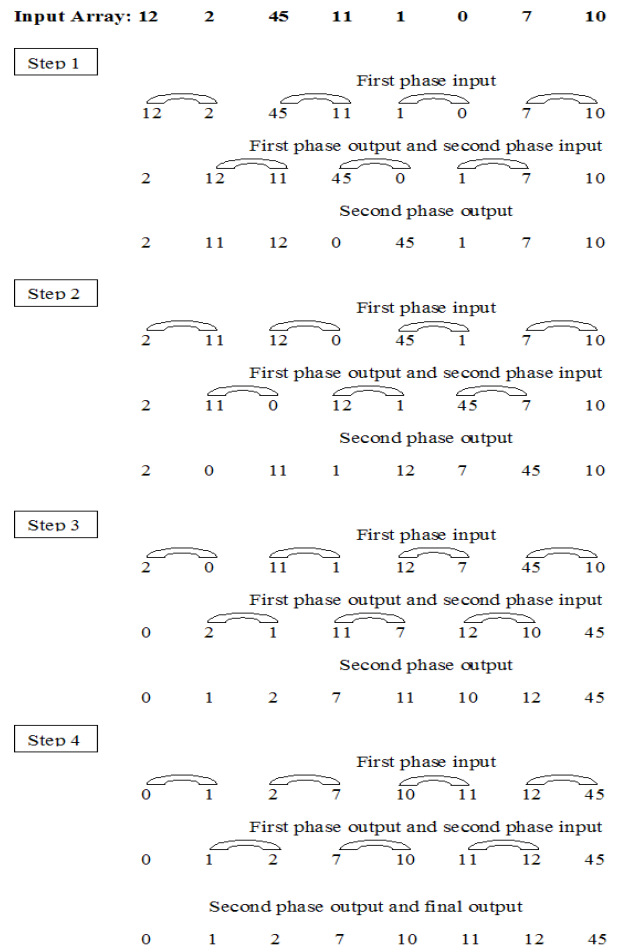


Figure 1. Example of sorting an array of 8 elements with the proposed Pair Wise Sorting Technique

### III. COMPLEXITY ANALYSIS OF PAIR WISE SORTING

The proposed technique has a special feature that a number of unnecessary passes can be cut down by using a flag variable. In this technique, the array can be sorted at any pass and not all time it is required to complete all of the passes. Flag variable is used for the purpose of reducing the total number of passes and therefore the total number of comparisons.

So the total number of comparisons varies in two different cases:

1. Without a Flag Variable
2. With a Flag Variable

#### A. Complexity Analysis: Without a Flag

In this case, the first check is the one that determines whether the number of elements in the array is even or odd and depending on this, the next step is chosen. It requires a comparison at the very beginning.

$$\text{So, } C_1 = 1 \quad (1)$$

Let, the number of elements in the array is  $n$

**When  $n$  is even:**

$$C_2 = p * q \quad (2)$$

$$\text{Here, } p = \text{total number of passes} = \frac{n}{2}$$

$$\begin{aligned} q &= \text{total number of comparisons at each pass} \\ &= \frac{n}{2} + \frac{n}{2} - 1 \\ &= n-1 \end{aligned}$$

Hence, from (2) we get

$$C_2 = \frac{n(n-1)}{2} \quad (3)$$

**When  $n$  is odd:**

For the last element, total comparisons will be,

$$\begin{aligned} C_3 &= \frac{(1+2+\dots+n-1)}{n-1} \\ &= \frac{\frac{n(n-1)}{2}}{n-1} \\ &= \frac{n}{2} \end{aligned} \quad (4)$$

A control variable  $k$  is used for this odd-even checking. Hence, from equation (1), (3) and (4) we get, total number of comparisons when no Flag variable is used is:

$$C = C_1 + C_2 + C_3 = 1 + \frac{(n-k)(n-k-1)}{2} + \frac{nk}{2}, \quad k = \begin{cases} 0, & n \text{ is even} \\ 1, & n \text{ is odd} \end{cases} \quad (5)$$

#### B. Complexity Analysis: With a Flag

In this case also, the first check is the one that determines whether the number of elements in the array is even or odd and depending on this, the next step is chosen. It requires a comparison at the very beginning.

$$\text{So, } C_1 = 1 \quad (6)$$

Let, the number of elements in the array is  $n$

**When  $n$  is even:**

$$\text{Number of comparisons in the first sub phase} = \frac{n}{2}$$

$$\text{Number of comparisons in the second sub phase} = \frac{n}{2} - 1; \text{ when } n > 1$$

After each phase, there is a check to detect whether any interchange has been made or not. Hence, the total number of comparisons after each phase is:

$$\frac{n}{2} + \left(\frac{n}{2} - 1\right) + 1; \text{ when } n > 1$$

$$\text{Therefore, the total number of comparisons is } = n + 2n + 3n + 4n + \dots + pn; \text{ where } p = \frac{n}{2}$$

$$= n(1 + 2 + 3 + \dots + p)$$

$$= n \cdot \frac{p(p+1)}{2}$$

$$= n \cdot \frac{\frac{n}{2} \cdot \left(\frac{n}{2} + 1\right)}{2}$$

$$= n^2 \left(\frac{n+2}{8}\right)$$

In average case, total number of comparisons,

$$\begin{aligned} C_2 &= \frac{\frac{n^2(n+2)}{8}}{\frac{n}{2}} \\ &= \frac{n(n+2)}{4} \end{aligned}$$

**When  $n$  is odd:**

For the last element, total comparisons will be,

$$C_3 = \frac{(1+2+\dots+n-1)}{n-1}$$

$$\begin{aligned} &= \frac{n(n-1)}{2} \\ &= \frac{n-1}{2} \end{aligned} \quad (8)$$

A control variable  $k$  is used for this odd-even checking. Hence, from equation (6), (7) and (8) we get, total number of comparisons when no Flag variable is used is:

$$C = C_1 + C_2 + C_3 = 1 + \frac{(n-k)(n-k+2)}{4} + \frac{nk}{2}, k = \begin{cases} 0, & n \text{ is even} \\ 1, & n \text{ is odd} \end{cases} \quad (9)$$

#### IV. COMPARISON WITH OTHER SORTING TECHNIQUES

The proposed technique has no major efficiencies over other sorting techniques rather it's running time is a little bit higher than other techniques. It is just a new method that sorts numeric data. Its symmetric structure makes it simpler than other sorting techniques. Table 1 shows a comparison of average case complexity for the proposed technique and other already existing sorting techniques:

TABLE I. COMPARISON TABLE OF DIFFERENT SORTING TECHNIQUES IN AVERAGE CASE COMPLEXITY

Sorting Technique	Average Case Complexity
Bubble Sort	$\frac{n(n-1)}{2} = O(n^2)$
Insertion Sort	$\frac{n(n-1)}{4} = O(n^2)$
Selection Sort	$\frac{n(n-1)}{2} = O(n^2)$
Quick Sort	$1.4 n \log n = O(n \log n)$
Merge Sort	$n \log n = O(n \log n)$
Heap Sort	$3n \log n = O(n \log n)$
Pair Wise Sort (Proposed Technique)	<p>Without a Flag:  <math>1 + \frac{(n-k)(n-k-1)}{2} + \frac{nk}{2} = O(n^2)</math></p> <p>With a Flag:  <math>1 + \frac{(n-k)(n-k+2)}{4} + \frac{nk}{2} = O(n^2)</math></p> <p>Here, <math>n</math> is the number of elements to be sorted  and <math>k = \begin{cases} 0, &amp; n \text{ is even} \\ 1, &amp; n \text{ is odd} \end{cases}</math></p>

We compared the performance of the proposed pair wise sorting technique with other sorting techniques namely Bubble Sort, Insertion Sort and Quick Sort for sorting large number of random numeric data. The comparison was based on time required by each method for sorting these data. Table 2 and Figure 2 show the result. The results shows that performance of

the proposed pair wise sorting technique is same as other sorting techniques for around 15000 data but it takes higher time with the increase in the number of elements to be sorted.

TABLE II. TIME REQUIRED BY EACH OF THE CANDIDATE METHODS FOR SORTING LARGE NUMBER OF RANDOM DATA

Number of Elements	Time in Milliseconds (ms)			
	Pair Wise Sort (Without Flag)	Bubble Sort	Insertion Sort	Quick Sort
5000	989.02	934.06	934.04	879.12
10000	2307.67	2142.86	2087.91	1868.13
15000	3846.13	3351.65	3351.65	2857.14
20000	5669.34	4780.22	4270.33	3846.15
25000	7637.36	6208.79	6208.79	4835.17

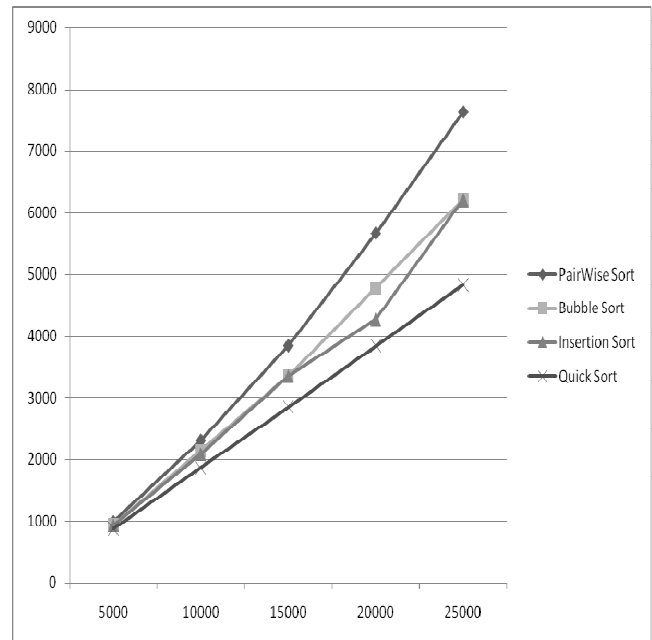


Figure 2. Performance graph

#### V. CONCLUSION AND FUTURE WORK

The proposed pair wise sorting technique is a simple one based on odd-even strategy. We strongly believe that this strategy can be further used to develop techniques which will be more efficient. However, the limitation of the proposed technique is, it requires different approach for the last member of the array when the array contains odd number of elements.

Most of the time, it is not required to complete all of the phases to sort elements. Here a flag is used to cut down the total number of phases. A flag is very useful for the cases when the elements of the array exist close to their exact positions. But the use of this flag causes the average number of comparisons to be higher than other algorithms. It would be best if it is possible to introduce a technique which will be able to sort elements efficiently without using a flag. This point will get our future focus.

## REFERENCES

- [1] D. E. Knuth, The Art of Computer Programming, Volume 3: Sorting and Searching, Second ed. Boston, MA: Addison-Wesley, 1998.
- [2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms, Second ed. MIT Press, Sep. 2001.
- [3] S. G. Akl, Parallel Sorting Algorithms. Orlando: Academic Press, Inc., 1990.
- [4] G. Graefe, "Implementing sorting in database systems," ACM Comput. Surv., vol. 38, no. 3, pp. 1–37, 2006.
- [5] C. Lauterbach, M. Garland, S. Sengupta, D. Luebke, and D. Manocha, "Fast BVH construction on GPUs," in Proc. Eurographics '09, Mar. 2009.
- [6] V. Havran, "Heuristic ray shooting algorithms," Ph.D. Thesis, Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Nov. 2000.
- [7] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.
- [8] B. He, W. Fang, N. K. Govindaraju, Q. Luo, and T. Wang, "Mars: A MapReduce framework on graphics processors," in Proc. 17th Int'l Conference on Parallel Architectures and Compilation Techniques, 2008, pp. 260–269.
- [9] Nadathur Satish, Mark Harris, and Michael Garland, "Designing efficient sorting algorithms for manycore GPUs", in Proc. 23rd IEEE International Parallel and Distributed Processing Symposium, May 2009, pp. 1-10.

## AUTHORS' PROFILE

Md. Jahangir Alam received his B.Sc. (Engg.) and M.Sc. in Computer Science and Engineering from Dhaka International University (DIU), Bangladesh. His research interests include Software Engineering and Data Structure. Currently he is working as a Senior Lecturer of Department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh.



Muhammad Monsur Uddin received his B.Sc. (Engg.) in Computer Science and Engineering from Institute of Science, Trade and Technology (ISTT), Dhaka, Bangladesh. His research interests include Software Engineering and Data Structure.

Mohammad Shabbir Hasan received his B.Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2008. His research interest includes different areas of Software Engineering like Requirement Engineering, Software Metric, Software Security and Software Maintenance. He has coauthored numerous research papers published in International Journals and Conference Proceedings. Currently he is working as a researcher of Panacea Research Lab, Dhaka, Bangladesh. He is also a lecturer of Department of Computer Science and Engineering in Institute of Science, Trade and Technology, Dhaka, Bangladesh.



Abdullah Al Mahmood received his B.Sc. (Engg.) in Computer Science and Engineering from Khulna University of Engineering and Technology (KUET), Bangladesh in 2009. His research interest includes Robotics, Artificial Intelligence, Internet Security and various areas of Software Engineering like Requirement Gathering, Requirement Prioritization and Software Security. He has coauthored a good number of research papers published in International Journals and Conference Proceedings. Currently he is working as a researcher of Panacea Research Lab, Dhaka, Bangladesh. He is also a lecturer at Institute of Science, Trade and Technology, Dhaka, Bangladesh and a Project Coordinator of Technocrats BD.



# Improving Client-Server Response Time Using IPv6 Header Suppresion Over MPLS

Imad J. Mohamad, Yung-Wey Chong  
National Advanced IPv6 Center  
Universiti Sains Malaysia  
11800 Penang, Malaysia  
[imad@nav6.org](mailto:imad@nav6.org), [chong@nav6.org](mailto:chong@nav6.org)

Tat-Chee Wan, Putra Sumari  
School of Computer Sciences  
Universiti Sains Malaysia  
11800 Penang, Malaysia  
[tcwan@cs.usm.my](mailto:tcwan@cs.usm.my), [putras@cs.usm.my](mailto:putras@cs.usm.my)

**Abstract**—Optimizing the response time for Client-Server IPv6 traffic over label switched path (LSP) is the main contribution for this paper. It is achieved using header suppression for real time IPv6 traffic across MPLS label switched path (LSP). Robust Header Compression (RoHC) and Payload Header Suppression (PHS) are two options defined in IEEE 802.16 for mobile WiMAX performance work using link-by-link approach. This paper adapts PHS for MPLS performance and extends it to work over LSP using end-to-end approach. The implementation for IPv6 header suppression using NS2 shows improvement in response time for client-server traffic by 1.7s. Additional improvement in QoS parameters for UDP and TCP traffic is investigated.

**Keywords**—component; Client-Server Traffic, LSP, IPv6, Header Suppression;

## I. INTRODUCTION

The motivation for this work is the increasing needs to investigate IPv6 performances with challenges of bigger header sizes compared to the smaller sizes of the packets payloads inside the internet backbone infrastructures like MPLS.

In computer networking, each protocol layer appends its own control information into a packet, forming a protocol header size before packet transmission over a network. This leads to increased overheads, especially for small data sizes. RoHC and PHS are two well known standards can be used in IEEE 802.16 wireless technology. Each scheme is developed to reduce upper layers overheads and to improve the number of supported users [1]. These header compression standards is used to compress significant amount of IP headers since it has lots of redundant fields. For example, an IPv6 stream has many redundant fields such as IP source and IP destination that causes a minimum of 32 bytes of redundancy of packet transmission overheads. In contrast, IPv4 has 8 bytes redundancy for the same fields.

Edge routers such MPLS provider edge (PE) of Internet backbone work as aggregation nodes in which high traffic load expected to pass through MPLS-LSP. Real-time and none real-time traffic contributes to the congestion state at Ingresses of MPLS. Client-server (HTTP/TCP) stream suffers from congestion and packets drop more than UDP because of priority if we ignore the packet size. Consequently TCP

retransmission and congestion increase the response time for client-server traffic.

In terms of QoS for real time traffic, the impact on VoIP applications are caused by different Service Level Agreement (SLA) metric parameters such as delay, throughput, jitter and packets drop [2]. QoS parameters for real time and for none real time traffic are investigated in this paper.

This paper is organized as follows: background information is provided in Section II. Related work is discussed in Section III. The proposed framework and simulation results are presented in Section IV.. Section V concludes this paper.

## II. BACKGROUNDS

### A. IPv6 over MPLS

The success of MPLS technology in providing QoS for real time IP applications makes it one of the favorite choices for ISPs when merged to IPv6 in Internet backbone networks. Therefore several IPv6 scenarios over MPLS have been identified in the literatures as a part of IPv6 deployments (Table 1) [3].

TABLE I. IPv6 OVER MPLS DEPLOYMENT SCENARIO [3]

Scenario	Impact on
IPv6 Tunnels configured on CE	No Impact on MPLS
IPv6 over Circuit_over_MPLS	No Impact on IPv6
IPv6 Provider Edge Router (6PE) over MPLS	No Impact on MPLS core
Native IPv6 MPLS	Require full network upgrade

### B. MPLS Technology

MPLS is a routing and forwarding protocol standardized by IETF in 2001[4]. Four main processes are known in dealing with entering packets into MPLS cloud. The first process is the classification process, in which ingress's incoming packets are classified and assigned to forwarding equivalent classes (FECs)



according to the required treatment similarity. The same MPLS label would be provided to all packets belonging to the same FEC. The second is the label push (or encapsulation) process, in which the MPLS label is pushed by ingress to prefix the packet header. The third process, forwarding, guides the encapsulated packet through an LSP using a label switching mechanism assisted by the label information base (LIB) table. The fourth is the final label pop (or decapsulation) process, which is maintained by egress (or penultimate) LSR, and followed by a return to normal layer 3 routing [5].

### C. Header Suppression verses Header Compression

Many header compression and suppression techniques are developed to conserve the bandwidth and to improve TCP/IP connectivity over long Round Trip Time (RTT) such as satellite (most costly), wireless (limited and costly) and high speed networks. Typically compression schemes use encoding, decoding and feedback-updating processes of packet header in order to synchronize their compressor state with the corresponding decompressor state. Suppression schemes operate by stripping out the header fields which are redundant in successive packets of a certain stream.

## III. RELATED WORK

Aref [6] discussed the importance of Internet QoS. A provider SLA is one of the important aspects of QoS. For example, the original SLA (before November 2007) between Ninsoft (www.ninsoft.com) and an Internet Service Provider (ISP) stated that, on a weekly basis, the ISP should ensure 98% of response time for some applications, so that email response time does not exceed 20 seconds, connection to hosted server does not exceed 1.5 seconds, and Website home page downloaded does not exceed 8 seconds. In existing work, improvement of HTTP response time for client-server Web traffic is one of the drivers to develop the proposed framework for MPLS.

RFC4247 [7] declared guidelines and requirements for header compression over MPLS and the necessity for applying header compression schemes to reduce header overhead. It considers MPLS usage in routing compressed packets over MPLS-LSP without compression/decompression cycles at each LSR that belongs to a certain LSP. Although applying hop-by-hop HC/HD over LSP decreases bandwidth requirements, it adds additional processing overhead in each router of a LSP. Compression of VoIP packets is one of concerns for RFC4247. It discussed the encapsulation of voice/RTP/UDP/IP with the addition of MPLS label to form 44 bytes header for /RTP/UDP/IPv4/MPLS-label encapsulation.

In Hadia paper [8], the effectiveness for a modified version of PHS is compared to PHS and RoHC of WiMAX. The outcome of this paper showed improvement in the number of users from 125 without header compression to 185, 250, and 260 using PHS, modified PHS, and RoHC respectively. The results show improvement for PHS using adaptive algorithm. The paper investigated real time traffic only.

In [9], RoHC and PHS approaches were compared in a certain case of VoIP transmission. It concluded that RoHC is better than PHS in terms of efficiency, and PHS efficiency may be better than ROHC when static headers size is more than none-static information.

Shekhar paper [10] discussed the support for heterogeneous traffic over MPLS backbone. It introduced distortion aware traffic engineering model to optimize the MPLS support for heterogeneous traffic. Our approach uses header suppression over MPLS-LSP to reduce the distortion (packets loss). In addition it investigates other QoS parameters such as delay, response time and throughput for heterogeneous traffic.

## IV. PROPOSED METHODOLOGY AND SIMULATION RESULTS

This paper is an extension of the algorithms and experiments in our previous work [11] for further IPv6 QoS investigation. It thoroughly investigates the effects of real time IPv6 header compression on none real time IPv6 traffic. The performance for client-server traffic is the major concern in current work.

The design, algorithms, justifications and considerations for the proposed approach are discussed in [11].

Simulation parameters and the topology for the experiment of this paper are shown in Table II and Figure 2, respectively. The HTTP experiment is implemented using a Web traffic generation model (PackMime-HTTP) of NS-2 [12]. The traffic intensity generated by PackMime-HTTP is controlled by the rate parameter, which is the average number of new connections initiated each second. It is capable of generating HTTP/1.0 and HTTP/1.1 (persistent, non-pipelined) connections.

The experiment investigates the effect of header compression when MPLS-LSP is shared between TCP/HTTP Web traffic of the client server and TCP/FTP and UDP/RTP traffic. PackMime-HTTP generates variable length packets for Web traffic because normally, the size of Web data received would depend on the client request type from the Web server. TCP/FTP and UDP/RTP packets are set up to fixed sizes of 1000 and 93 bytes, respectively.

Figure 3 shows that the implementation of LSP-PHS reduces the packet drop for UDP and retransmission of TCP packets for HTTP and FTP traffic. The packet drop for UDP data declines from 50 to 0. This is because the bandwidth allocation is given to UDP before HTTP and TCP/FTP. When LSP-PHS is implemented, HTTP packet drop declines from 294 to 248 or 16%, whereas TCP/FTP packets decline from 32 to 17 or 53%. TCP/HTTP streams have the highest packet drop compared with UDP and TCP/FTP. This is because packet drops occur between the HTTP server and the client, which cannot be reduced by the implementation of LSP-PHS.

TABLE II. SIMULATION PARAMETERS

VoIPv6 Parameters	Setting	FTP and HTTP Parameters	Setting
Flows (0 –4)	IPv6/UDP/RTP/Voice	type of TCP0 type of TCP1	FTP/TCP/IPv6 HTTP/TCP/IPv6
Codec	GSM 6.0	HTTP connection rate	2 connection per second
Voice payload size	33 bytes	HTTP module	PackMime-HTTP
Total Pkt Size	93 bytes	TCP0 Pkt Size TCP1 Pkt Size	1000 bytes Variable length
Transmission rate	13Kbps CBR	-	-
Source nodes	0,1,2,3 and 4	TCP0 Source node TCP1 Source node	5 6
Destination nodes	13,14,15, 16 and 17	TCP0 sink TCP1 sink	18 19

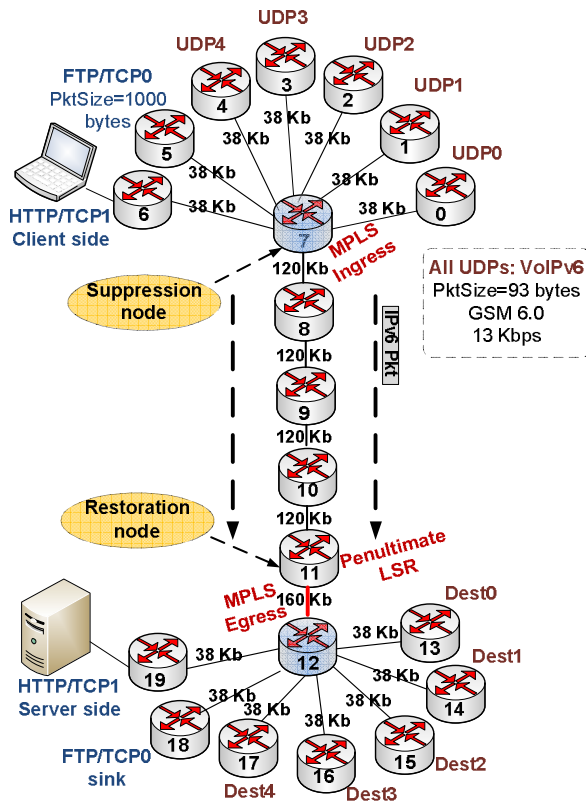


Figure 2. MPLS topology (Scenario)

Figure 4 shows an improvement in TCP/FTP throughput when LSP-PHS is implemented. The HTTP response time also shows marked improvement by 1730 ms when the MPLS cloud supported LSP-PHS (Figure 5). The improvement value appears more prominently in the response time than in the reduction of packet drop for HTTP.

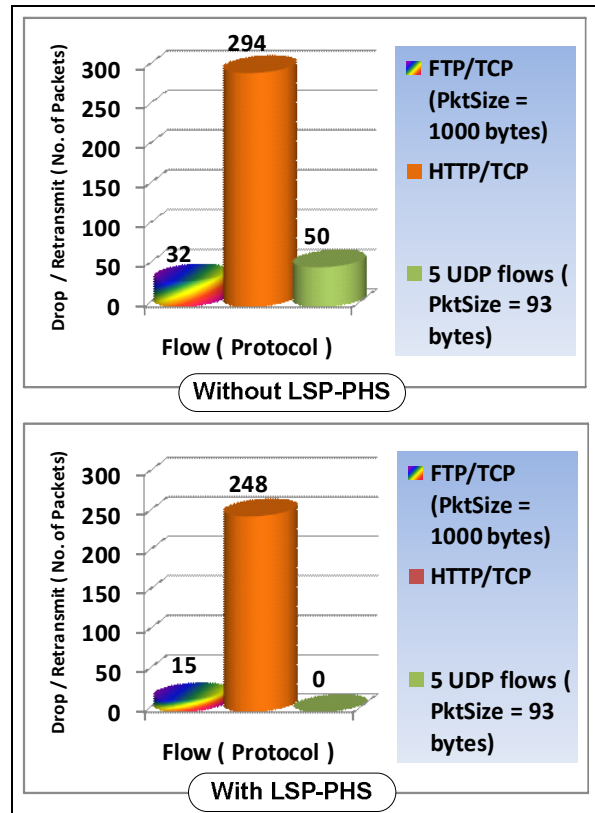


Figure 3. Drop/Retransmit for UDP/TCP Statistics

Figures 6, 7, and 8 show improvements in UDP throughput, and delay; when LSP-PHS is implemented, the delay metric for VoIPv6 UDP streams is reduced to the half on average. Throughput metric is improved and better tuned. Only minor tuning is observed in terms of jitter for UDP streams. Table III summarizes the results of the experiment.

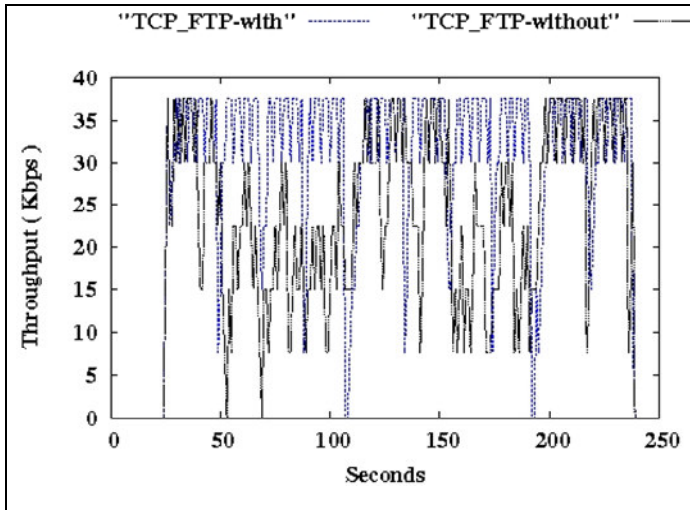


Figure 4. FTP/TCP Throughput (With and Without) LSP-PHS

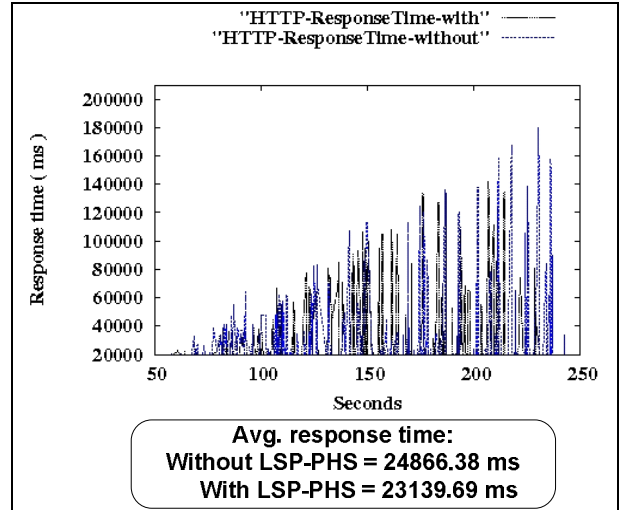


Figure 5. Response time for HTTP/TCP (With and Without) LSP-PHS

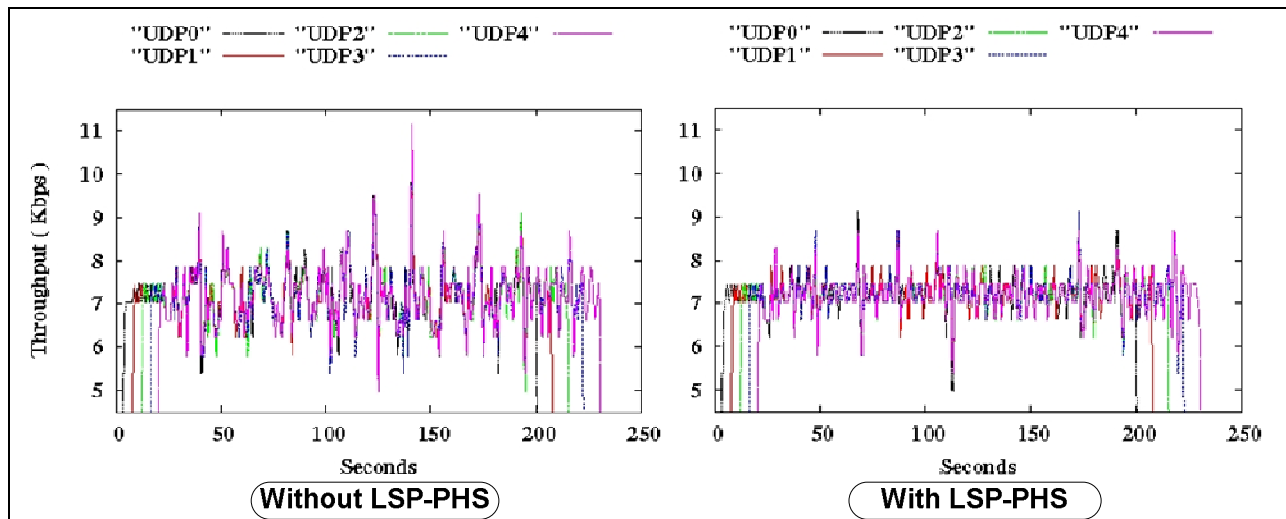


Figure 6. UDP Throughput

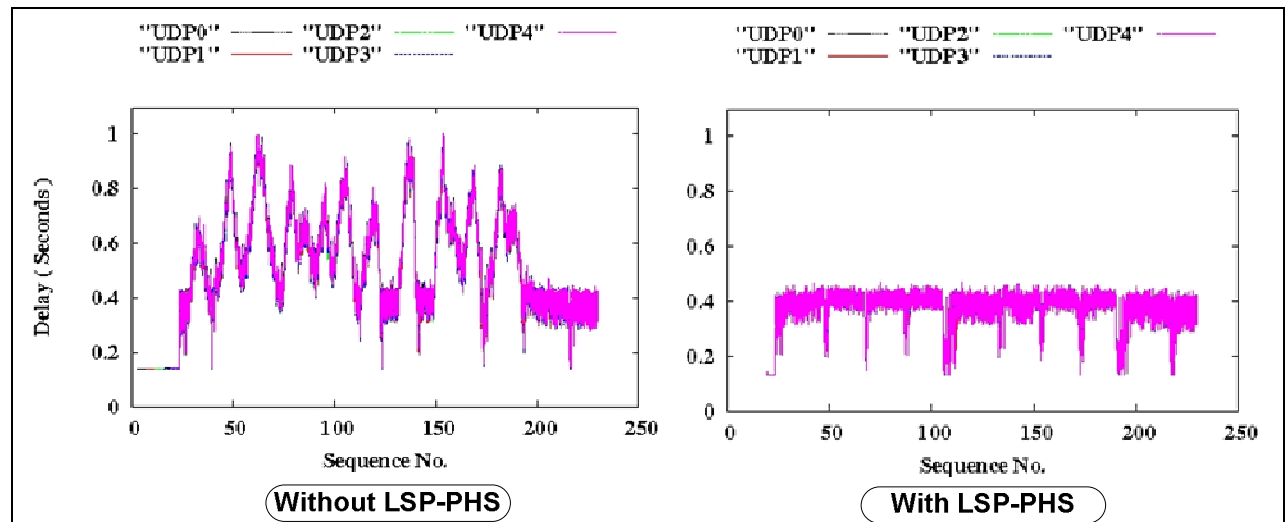


Figure 7. UDP Delay

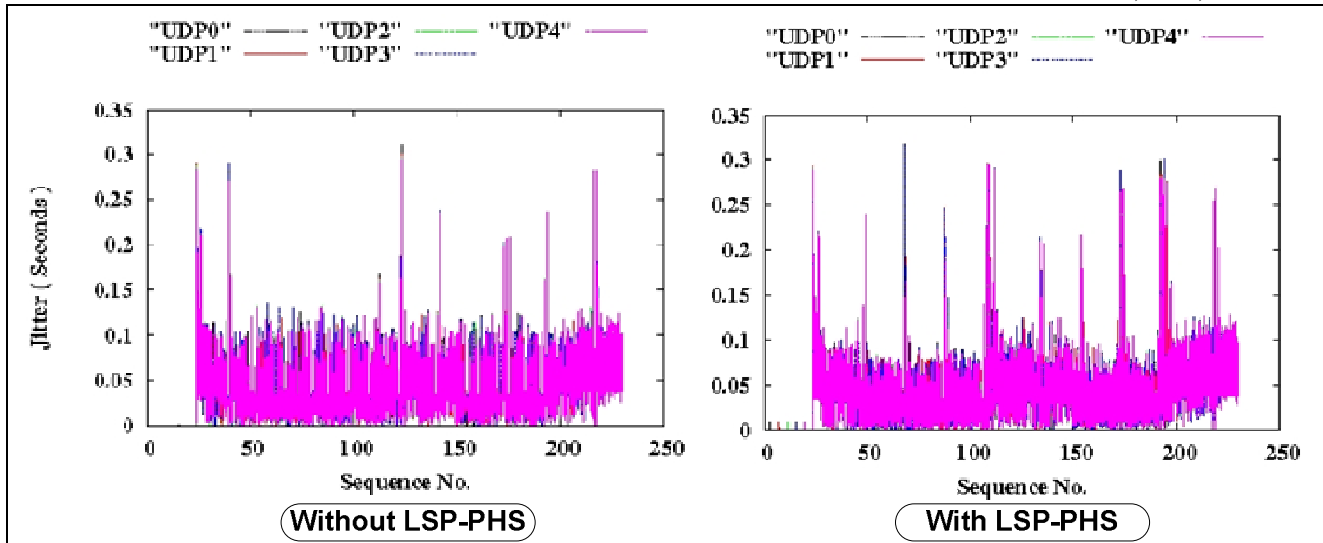


Figure 8. UDP Jitter (Delay variation)

TABLE III. SUMMARY FOR RESULTS

QoS metrics	UDP (93 bytes)		FTP/TCP0 (1000 bytes)Pkt		HTTP/TCP1 (variable length) Pkt	
	Without LSP-PHS	With LSP-PHS	Without LSP-PHS	With LSP-PHS	Without LSP-PHS	With LSP-PHS
Throughput	High variation, Btw 5.5 to 11.5 Kbps	Tuned with Less variation Btw 5.5 to 9.5 Kbps	High variation and most time Btw 8 to 30 Kbps	Maximized and Tuned up Btw 30 to 38 Kbps	N/A	
Delay	high variation 0.2 to 1 sec	Reduced to 50%, Btw 0.2 to 0.5 sec	N/A		N/A	
Jitter	Minor improvement		N/A		N/A	
UDPdrop or TCPPretransmit	50 Pkt	Reduced to 0 Pkt Improved 100%	32 Pkt	15 Pkt, Improved 53%	294 Pkt	248 Pkt, Improved 16%

## V. CONCLUSION

In terms of end-to-end QoS metrics, LSP-PHS implementation using NS2 shows considerable reduction of approximately 50% in UDP maximum delay. This reduction results in diminished delay response with a faster response time of around 1.7 seconds in HTTP and better tuned UDP throughput. It also provides minor improvement in UDP jitter. In addition, LSP-PHS implementation also eliminates packet drop for real-time traffic (VoiPv6).

## REFERENCES

- [1] So-In, C., R. Jain, and A.-K. Tamimi, Capacity Evaluation for IEEE 802.16e Mobile WiMAX. Journal of Computer Systems, Networks, and Communications, 2010.
- [2] Evans, J.W. and C. Filsfil, Deploying IP and MPLS QoS for Multiservice Networks: Theory & Practice. Morgan Kaufmann series in networking. 2007: ELSEVIER.
- [3] Grivaud, J.-C. Cisco IPv6 Solutions, Integration & Co-Existence. 2008; Available from: [http://cco.cisco.com/web/SI/ciscoexpo08/presentations/IPv6\\_re\\_itve\\_Jean\\_Charles\\_Grivaud.pdf](http://cco.cisco.com/web/SI/ciscoexpo08/presentations/IPv6_re_itve_Jean_Charles_Grivaud.pdf).
- [4] Rosen, E., A. Viswanathan, and R. Callon, Multiprotocol Label Switching Architecture. 2001.

- [5] Mine, I. and J. Lucek, MPLS-Enabled Applications: Emerging Developments and New Technologies, ed. S.o.C.N.D. Systems. 2008: John Wiley & Sons
- [6] Meddeb, A., Internet QoS: Pieces of the puzzle. Communications Magazine, IEEE, 2010. 48(1): p. 86-94.
- [7] J. Ash, e.a. Requirements for Header Compression over MPLS. [RFC4247] 2005.
- [8] El-Hennawy, H.M., H.H. Gomaa, and A.S. Amin. The effectiveness of transmitting voice traffic over IPv6 convergence sublayer of WiMAX mesh networks. in Telecommunications (ICT), 2010 IEEE 17th International Conference on. 2010.
- [9] Nuaymi, L., et al. Headers Overhead Estimation, Header Suppression and Header Compression in WiMAX. in Wireless and Mobile Computing, Networking and Communications, 2007. WiMOB 2007. Third IEEE International Conference on. 2007.
- [10] Srivastava, S., A.v.d. Liefvoort, and D. Medhi, Traffic engineering of MPLS backbone networks in the presence of heterogeneous streams. Comput. Netw., 2009. 53(15): p. 2688-2702.
- [11] Mohamad, I.J., et al. Optimizing the MPLS support for real time IPv6-Flows using MPLS-PHS approach. in TENCON 2009 - 2009 IEEE Region 10 Conference. 2009.
- [12] Jin, C., et al. Stochastic models for generating synthetic HTTP source traffic. in INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies. 2004.



# Hybrid Compression of Color Images with Larger Trivial Background by Histogram Segmentation

M.Mohamed Sathik<sup>1</sup>, K.Senthamarai Kannan<sup>2</sup> and Y.Jacob Vetha Raj<sup>3</sup>

<sup>1</sup>Department of Computer Science, Sadakathullah Appa College, Tirunelveli, India  
mmdsadiq@gmail.com

<sup>2,3</sup>Department of Statistics, Manonmanium Sundaranar University, Tirunelveli, India  
senkannan2002@gmail.com  
jacobvetharaj@gmail.com

**Abstract-** A hybrid image compression method is proposed by which the background of the image is compressed using lossy compression and the rest of the image is compressed using lossless compression. In Hybrid Compression of Color Images with Larger Trivial Background by Histogram Segmentation(HCCILTBHS), input color image is subjected to binary segmentation using histogram to detect the background. The color image is compressed by standard lossy compression method. The difference between the lossy image and the original image is computed and is called as residue. The residue at the background area is dropped and rest of the area is compressed by standard lossless compression method. This method gives lower bit rate than the lossless compression methods and is well suited to any color image with larger trivial background.

**Key Words-** Segmentation, Erosion, Dilation, Image Compression.

## I. INTRODUCTION

Even though the memory capacities of computers have increased as new technologies are emerging, the requirement of more storage space is also increasing as more data are needed to be stored. In the case of image data, the spatial and color resolutions are increased for the betterment of image quality, thus requires more space to store images. Image compression is one of the solutions to meet the storage requirements. In image compression, there are two major classifications; they are lossless [1]-[5] and lossy [6] compression. In lossless image compression, the entire data can be restored after decompression, but not in the case of lossy compression. Vector quantization [7],[8], wavelet transformation [9]-[13] techniques are widely used in addition to various other methods[15]-[17] in image compression. The problem in lossless compression is that, the compression ratio is very less; where as in the lossy compression the compression ratio is very high but may lose vital information of the image. The quality of images are measured in lossy compression methods using various techniques [18],[19]. Some of the works carried out in hybrid image compression [20],[21] incorporated different compression schemes like PVQ and DCTVQ in a single image compression. But the proposed method uses lossless and lossy compression methods like [22] to compress a single image

The proposed method performs a hybrid compression, which makes a balance on compression ratio and image quality by preserving the vital information. In this approach the main subject in the image is very important than the background image. Considering the importance of image components, and the effect of

smoothness in image compression, this method segments the image as main subject and background, then the background of the image is subjected to lossy compression and the main subject is kept unaffected.

In the proposed work, for image compression, segmentation and morphological operations [23] are used. For segmentation, morphological operations such as erosion, dilation and closing there are lots of work has been carried out [24], [25]. A simple and a time efficient method for segmentation used in the proposed work is described in section II, section III gives a detailed description of the proposed method, the results and discussion are given in section IV and the concluding remarks are given in section V.

## II SEGMENTATION USING HISTOGRAM AND RUN LENGTH ENCODEING

Let  $X$  be a matrix of order  $m \times n \times p$ , represents the color image of width  $m$  and height  $n$ . Here  $p$  represents the number of color planes. For RGB color images the value of  $p$  is 3. The domain for  $X_{i,j,k}$  is  $[0..255]$ , for any  $i=1..m$ , any  $j=1..n$  and any  $k=1..p$ .

### A) Segmentation Using Histogram

The architecture of segmentation using histogram is shown in figure 1. To make the process faster the high resolution input color image is down sampled 2 times. When the image is down sampled each time the dimension is reduced by half of the original dimension. So the final down sampled image (D) is of the dimension  $\frac{m}{4} \times \frac{n}{4} \times p$ .

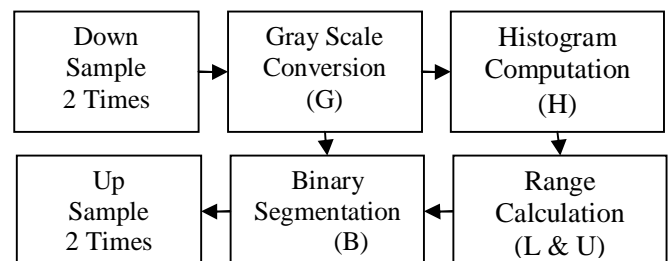


Figure -1

The down sampled color image is converted to gray scale image using equation (1).

$$G_{i,j} = 0.299 D_{i,j,0} + 0.587 D_{i,j,1} + 0.114 D_{i,j,2} \dots (1)$$

The histogram (H) is computed for the gray scale image. The most frequently present gray scale value ( $M_h$ ) is determined from the histogram as indicated by a vertical line in figure 2 by equation (2).

$$Mh = \arg\{\max(H(x))\} \quad \dots(2)$$

The background value of the images is having the highest frequency in the case of homogenous background. In order to surmise background textures a range of gray level

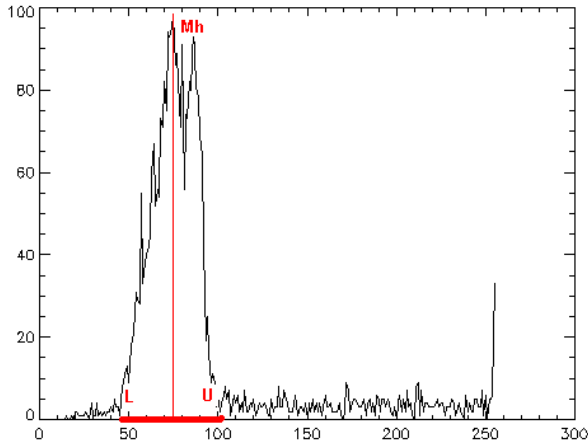


Figure - 2

values are considered for segmentation. The range is computed using the equations (3) and (4).

$$L = \max(Mh - 30, 0) \quad \dots(3)$$

$$U = \min(Mh + 30, 255) \quad \dots(4)$$

The gray scale image  $G$  is segmented to detect the background area of the image using the function given in equation (5)

$$B_{ij} = (G_{ij} > L) \text{ and } (G_{ij} < U) \quad \dots(5)$$

After processing the pixel values for background area is 1 in the binary image  $B$ . To avoid the problem of over segmentation the binary image is subjected to sequence of morphological operations. The binary image is eroded with smaller circular structural element (SE) to remove smaller segments as given in equation (6).

$$B = B \odot SE \quad \dots(6)$$

Then the resultant image is subjected to morphological closing operation with larger circular structural element as given in equation (7).

$$B = B \bullet SE \quad \dots(7)$$

### B) Run Length Encoding (RLE)

The run length encoding method is used to compress the binary segmented image. The number of continuous zeros and ones available alternatively in every line of the image counted and are stored. The decoding can be achieved by reading the sequence of numbers and reproduce the binary image by placing zeros and ones accordingly.

## III. PROPOSED HYBRID IMAGE COMPRESSION METHOD

The 'Hybrid Compression of High Resolution Color Images with Larger Trivial Background Images (HCCILTBHS)' method compresses the image with insignificant loss in background of the image and with no

loss in the vital area of the image. To achieve this, HCCILTBHS first converts the input RGB image into Gray Scale image ( $G$ ) and segments it into background and foreground image as described in section II. The segmented image ( $\beta$ ) is a binary image determined by a histogram based algorithm as described in section II A. The lossy compressed RGB color image ( $\lambda$ ) is created by compressing the original input image by standard lossy compression method. Then a residual image ( $\delta$ ) is computed by finding the difference between original image and lossy compressed image at vital area and is compressed by standard lossless compression method. The  $\lambda$ ,  $\delta_c$  and  $\beta_c$  are stored which are used to formulate the hybrid image ( $\hat{h}$ ). The residue at foreground area is a linear vector of size  $K$  and is extracted as per equation (8)

$$\delta_{ik} = \alpha_{ij} - \lambda_{ij} \text{ if } \beta_{ij} \text{ is } 1 \quad \dots(8)$$

Finally the composed  $\hat{h}$  is stored to get the Compressed Image. The compression ratio is expected to be comparatively better than lossless image compression and the image quality is to be better than lossy compression. More details are discussed in section IV. The block diagram in figure 3 gives the overall picture of this method. The entire operation in HCCILTBHS can be written in steps as follows

### HCCILTBHS Algorithm:

1. Read the RGB input image  $\alpha$ .
2. Down Sample  $\alpha$  two times to get  $D$ .
3. Convert  $D$  into Gray Scale Image  $G$ .
4. Segment to detect background area ( $\xi$ ) from  $G$ .
5. Up sample  $\xi$  two times to get  $\beta$ .
6. Lossy compress  $\alpha$  to get  $\lambda$ .
7. Find the residual image  $\delta$  by subtracting  $\lambda$  from  $\alpha$  at foreground area with regard to  $\beta$ .
8. Lossless compress  $\delta$  to get  $\delta_c$ .
9. Lossless compress  $\beta$  by RLE to get  $\beta_c$ .
10. Store  $\beta_c$ ,  $\lambda$  and  $\delta_c$ .

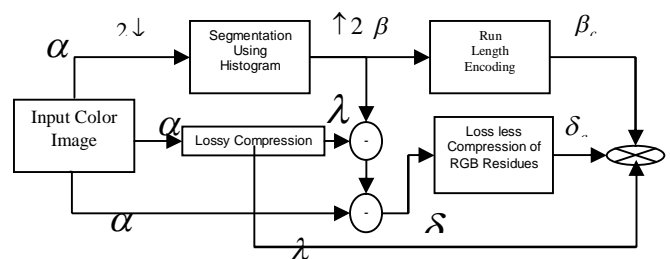


Figure – 3 Block Diagram of HCCILTBHS

The decompression is the reverse of the above operations. Initially all the compressed components  $\lambda$ ,  $\beta_c$  and  $\delta_c$  are restored from compressed file. The lossy image  $\lambda$  is decompressed; the binary segment information is loaded by run length decoding method. The lossless

information on residuals can be restored from  $\delta_c$ . The hybrid image can be formed by adding the residue  $\delta$  at vital area with the lossy image.

#### IV RESULTS AND DISCUSSION

The HCCILTBHS method is implemented according to the description in section III and tested with a set of twelve images shown in figure 5. The results obtained from the implementation of the proposed algorithms are shown in figures from 4 ,6 and table I. The compressed hybrid image by HCCILTBHS is shown in figure 4.e. Figure

4.a shows the original input image, figure 4.c shows the lossy compressed image of 4.a. In Figure 4.b the white area shows the detected vital area. From figure 4.d the error between lossy compressed image and the input image. It can be observed from figure 4.f that there is no difference in hybrid image from input image at foreground area. This indicates that there is no loss at foreground area. There are small differences in the background area can be observed. The compressed bit rates of the twelve test images are computed and tabulated in table 1.

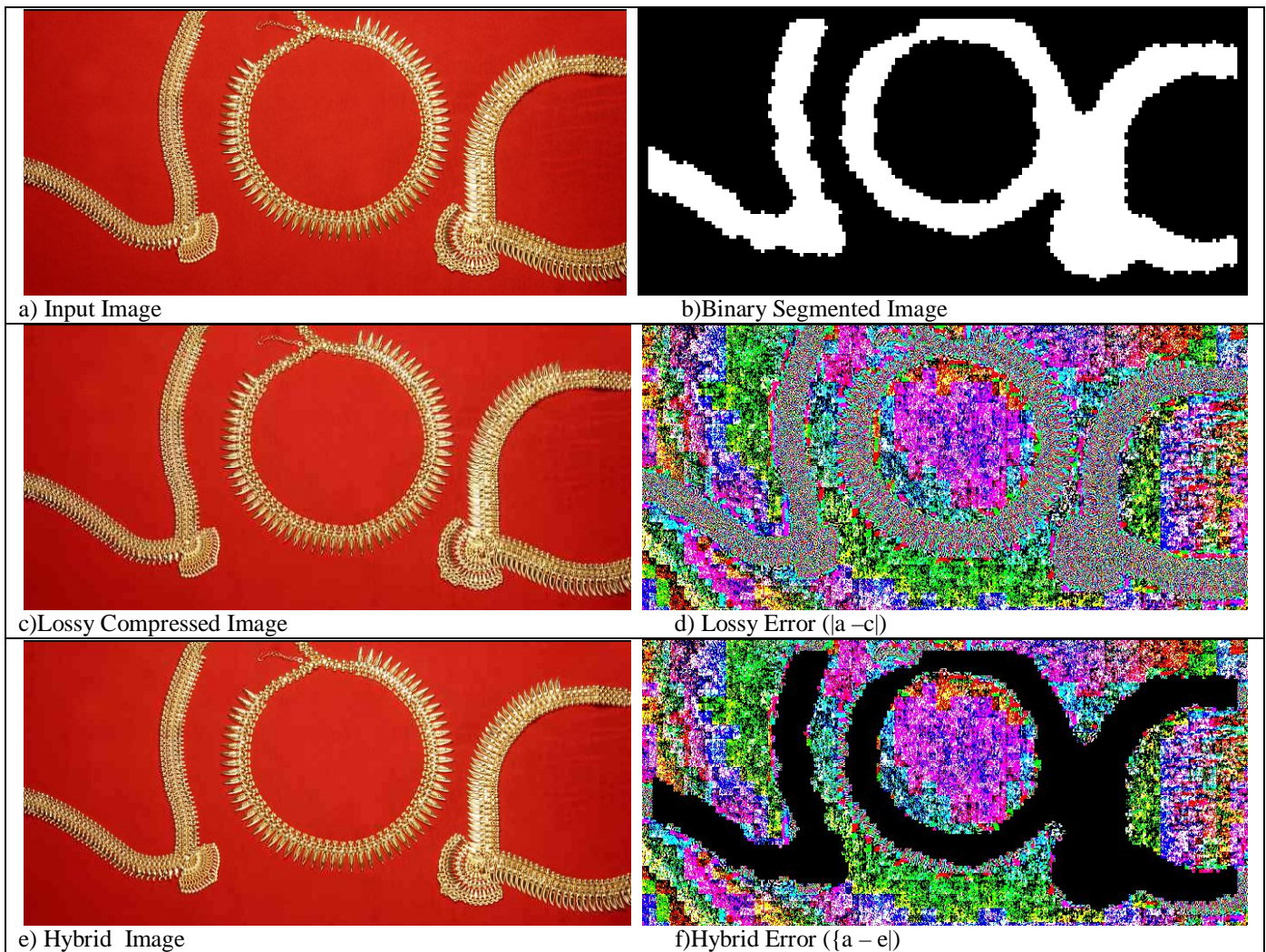


Figure 4– Input and output images.

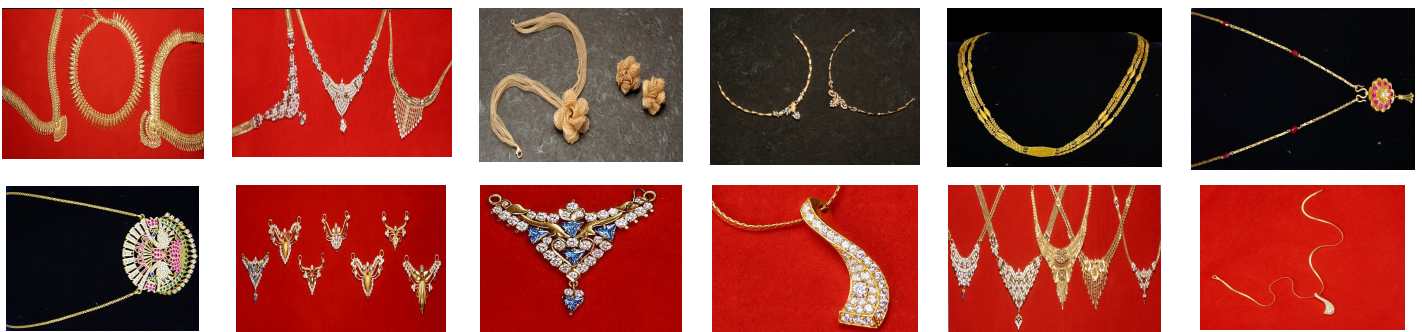


Figure 5 - Test Images (1 to 12 from left to right and top to bottom)



Table I – Bits Per Pixel (bpp) for Lossless, Lossy and Hybrid Compression Methods for test Images.

Images	Loss Less Bit Rate	Lossy Bit Rate	Hybrid Bit rate
1	12.4661	0.6181	10.5004
2	12.4370	0.3663	4.6582
3	14.082	0.8794	13.241
4	13.5499	0.5938	3.6106
5	5.9264	0.3557	5.7787
6	7.9622	0.3481	5.5602
7	10.8759	0.6363	9.6273
8	11.8418	0.5843	9.9861
9	14.4508	0.8725	11.5291
10	14.4690	0.6516	6.6611
11	7.2551	0.2938	1.6436
12	7.4225	0.4597	4.9224

The 'lossless bit rate' column indicates the bit rates achieved by compressing the input image by standard (JPEG200) compression method in lossless mode. The 'Lossy Bit Rate' column of table 1 indicates the bit rates achieved by standard JPEG compression method in lossy mode (Quality 20). The 'Hybrid Bit rate' column indicates the bit rate achieved by the proposed method (HCCILTBHS). Figure-6 shows the bit rates of compressed test images using lossless, lossy and proposed method. It can be easily noticed that the bit rate achieved by HCCILTBHS is less comparing to lossless bit rate and is higher than the bit rate of lossy compression bit rate.

## V. CONCLUSION

In HCCILTBHS, the bit rate is lesser than lossless compression and quality is higher than lossy compression methods. The computation time is higher because of the overhead of segmenting background and storing this information and residual. The bit rate depends on the percentage of foreground area of the image. As the area increases the bit rate also increases since the numbers of pixels to be stored in lossless mode are increased. Since the foreground of the image is preserved, this compression method can be well suited for any kind of image database which compresses images with larger and trivial background in offline. Improved segmentation and lossless compression methods may be incorporated in future to get better results.

## ACKNOWLEDGEMENT

The authors express their gratitude to University Grant Commission and Manonmanium Sundaranar University for financial assistance under the Faculty Development Program.

## REFERENCES

- [1] Xiwen OwenZhao, Zhihai HenryHe, "Lossless Image Compression Using Super-Spatial Structure Prediction", IEEE Signal Processing Letters, vol. 17, no. 4, April 2010 383

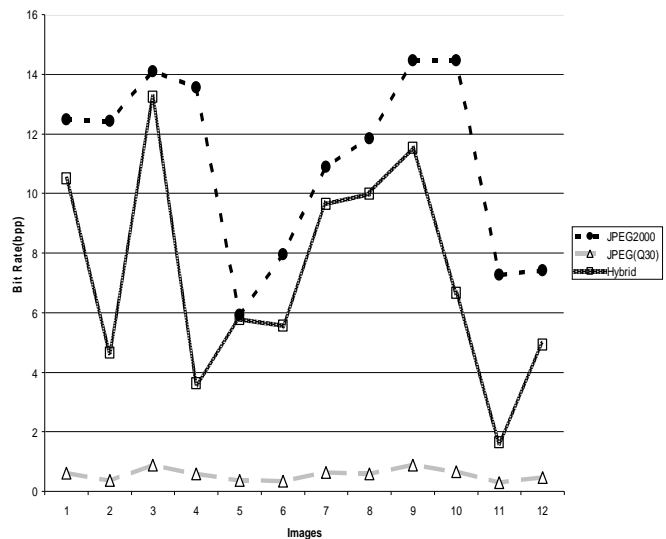


Figure – 6 Bit rates(bpp) obtained by Lossless, Lossy and Hybrid Compression Methods for test Images

- [2] Aaron T. Deever and Sheila S. Hemami, "Lossless Image Compression With Projection-Based and Adaptive Reversible Integer Wavelet Transforms", *IEEE Transactions on Image Processing*, vol. 12, NO. 5, May 2003.
- [3] Nikolaos V. Boulgouris, Dimitrios Tzovaras, and Michael Gerassimos Strintzis, "Lossless Image Compression Based on Optimal Prediction, Adaptive Lifting, and Conditional Arithmetic Coding", *IEEE Transactions on Image Processing*, vol. 10, NO. 1, Jan 2001
- [4] Xin Li, and Michael T. Orchard "Edge-Directed Prediction for Lossless Compression of Natural Images", *IEEE Transactions on Image Processing*, vol. 10, NO. 6, Jun 2001.
- [5] Jaemoon Kim, Jungsoo Kim and Chong-Min Kyung, "A Lossless Embedded Compression Algorithm for High Definition Video Coding" 978-1-4244-4291 / 09 2009 IEEE. ICME 2009
- [6] Rene J. van der Vleuten, Richard P. Kleihorst, Christian Hentschel, "Low-Complexity Scalable DCT Image Compression", *2000 IEEE*.
- [7] K. Somasundaram, and S. Domnic, "Modified Vector Quantization Method for Image Compression", *Transactions On Engineering, Computing And Technology Vol 13* May 2006
- [8] Mohamed A. El-Sharkawy, Chstian A. White and Harry, "Subband Image Compression Using Wavelet Transform And Vector Quantization", *1997 IEEE*.
- [9] Amir Averbuch, Danny Lazar, and Moshe Israeli, "Image Compression Using Wavelet Transform and Multiresolution Decomposition", *IEEE Transactions on Image Processing*, vol. 5, NO. 1, Jan 1996.
- [10] Jianyu Lin, Mark J. T. Smith, "New Perspectives and Improvements on the Symmetric Extension Filter Bank for Subband /Wavelet Image Compression", *IEEE Transactions on Image Processing*, vol. 17, NO. 2, Feb 2008.
- [11] Yu Liu, Student Member, and King Ngi Ngan, "Weighted Adaptive Lifting-Based Wavelet Transform for Image Coding", *IEEE Transactions on Image Processing*, vol. 17, NO. 4, Apr 2008.
- [12] Michael B. Martin and Amy E. Bell, "New Image Compression Techniques Using Multiwavelets and Multiwavelet Packets", *IEEE Transactions on Image Processing*, vol. 10, NO. 4, Apr 2001.
- [13] Roger L. Claypoole, Jr, Geoffrey M. Davis, Wim Sweldens, "Nonlinear Wavelet Transforms for Image Coding via Lifting", *IEEE Transactions on Image Processing*, vol. 12, NO. 12, Dec 2003.
- [14] David Salomon, "Data Compression, Complete Reference", *Springer-Verlag New York, Inc.* ISBN 0-387-40697-2.
- [15] Eddie Batista de Lima Filho, Eduardo A. B. da Silva Murilo Bresciani de Carvalho, and Frederico Silva Pinag  "Universal

- Image Compression Using Multiscale Recurrent Patterns With Adaptive Probability Model” , *IEEE Transactions on Image Processing*, vol. 17, NO. 4, Apr 2008.
- [16] Ingo Bauermann, and Eckehard Steinbach, “RDTC Optimized Compression of Image-Based Scene Representations (Part I): Modeling and Theoretical Analysis” , *IEEE Transactions on Image Processing*, vol. 17, NO. 5, May 2008.
- [17] Roman Kazinnik, Shai Dekel, and Nira Dyn , “Low Bit-Rate Image Coding Using Adaptive Geometric Piecewise Polynomial Approximation”, *IEEE Transactions on Image Processing*, vol. 16, NO. 9, Sep 2007.
- [18] Marta Mrak, Sonja Grgic, and Mislav Grgic ,”Picture Quality Measures in Image Compression Systems”, *EUROCON 2003 Ljubljana, Slovenia*, 0-7803-7763-W03 2003 IEEE.
- [19] Alan C. Brooks, Xiaonan Zhao, , Thrasyvoulos N. Pappas., “Structural Similarity Quality Metrics in a Coding Context: Exploring the Space of Realistic Distortions”, *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1261–1273, Aug 2008.
- [20] Hong, S. W. Bao, P., “Hybrid image compression model based on subband coding and edge-preserving regularization”, *Vision, Image and Signal Processing, IEE Proceedings*, Volume: 147, Issue: 1, 16-22, Feb 2000
- [21] Zhe-Ming Lu, Hui Pei ,”Hybrid Image Compression Scheme Based on PVQ and DCTVQ “,*IEICE - Transactions on Information and Systems archive*, Vol E88-D , Issue 10 ,October 2006
- [22] Y.Jacob Vetha Raj, M.Mohamed Sathik and K.Senthamarai Kanna,, “Hybrid Image Compression by Blurring Background and Non-Edges. *The International Journal on Multimedia and its applications*.Vol 2, No.1, February 2010, pp 32-41
- [23] William K. Pratt ,”Digital Image Processing” ,*John Wiley & Sons, Inc*, ISBN 9-814-12620-9.
- [24] Jundi Ding, Runing Ma, and Songcan Chen,”A Scale-Based Connected Coherence Tree Algorithm for Image Segmentation”, *IEEE Transactions on Image Processing*, vol. 17, NO. 2, Feb 2008.
- [25] Kyungsuk (Peter) Pyun, , Johan Lim, Chee Sun Won, and Robert M. Gray, “Image Segmentation Using Hidden Markov Gauss Mixture Models”, ” ,*IEEE Transactions on Image Processing*, VOL. 16, NO. 7, JULY 2007



# Realization and Study of High Performance Voltage Mode Oscillator based on CCCCTA: A Building Block for Analog Signal Processing

Deependra Pandey<sup>1</sup> and Prof.(Dr.) L.K. Singh<sup>2</sup>

<sup>1</sup> Asst. Professor, Dept. of ECE, Amity University, Lucknow

<sup>2</sup> Director, IET, Dr. R. M. L. Avadh University, Faizabad

## ABSTRACT

*At present there is a growing interest in designing current mode circuits. This attributed to their large signal bandwidth, great linearity, wider dynamic range, simple circuitry and low power consumption. The paper presents a basic current-mode building block for analog signal processing, namely current controlled current conveyor transconductance amplifier (CCCCTA). Its parasitic resistance at current input port can be controlled by an input bias current. It is very suitable to use in a current-mode signal processing, which is continually more popular than a voltage one. The proposed element is realized in a CMOS technology and is examined the performances through PSPICE simulations. The CCCCTA performs tuning over a wide current range. In addition, some circuits for example as a current-mode universal biquad filter and a grounded inductance occupy only single CCCCTA.*

**Keywords:** Current Conveyors, CCCCTA, Current-mode circuits, Voltage Mode Oscillator

## 1 INTRODUCTION

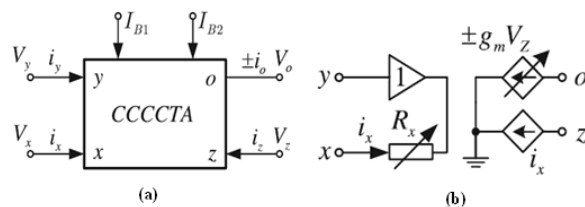
In the last decade, there has been much effort to reduce the supply voltage of electronic circuits. This is due to the command for portable and battery-powered equipments. Since a low-voltage operating circuit becomes necessary, the current-mode technique is ideally suited for this purpose more than the voltage-mode one. Current controlled current conveyor transconductance amplifier (CCCCTA) is the modified version of current conveyor transconductance amplifier (CCTA). CCTA, seems to be a versatile component in the realization of a class of analog signal processing circuits, especially analog frequency filters. It is seen that the CCTA cannot be controlled by the parasitic resistance at the input port, so when it is used in some circuits, it must unavoidably require some external passive components, especially resistors. This makes it inappropriate for IC implementation, due to

the requirement more chip area. In addition, the mentioned CCTA has a third-generation current conveyor (CCIII) as its input stage which has less flexibility for applications than a second-generation current conveyor (CCII). But the parasitic resistance at the input current port can be controlled by an input bias current in CCCCTA. Even though CCCCTA is implemented by employing CCCII and OTA, it is very convenient and useful if the CCCCTA is realized in monolithic chip to compact the circuits and systems. The performances of the proposed CCCCTA are illustrated by PSPICE simulations, and they show good agreement with the analysis.

## 2 BASIC CONCEPT OF CCCCTA

A5-terminals active element, namely current conveyor transconductance amplifier (CCTA), seems to be a versatile component in the realization of a class of analog signal processing circuits.

However, it is seen that the CCTA can not be controlled the parasitic resistance at input port so when it is used in some circuits, it must unavoidably require some external passive components, especially the resistors. This makes it not appropriate for IC implementation due to occupying more chip area. The purpose of this paper is to design and study a modified-version CCTA, which is newly named current controlled current conveyor transconductance amplifier (CCCCTA). The parasitic resistance at current input port can be controlled by an input bias current, and then it does not need a resistor in practical applications. The performances of proposed CCCCTA are illustrated by PSPICE simulations, they show good agreement as mentioned. CCCCTA properties are similar to the conventional CCTA, except that the CCCCTA has finite input resistance  $R_x$  at the X-input terminal. This parasitic resistance can be controlled by the bias current  $I_{B1}$ . The symbol and the equivalent circuit of the CCCCTA are illustrated in Fig.1(a) and (b), respectively.

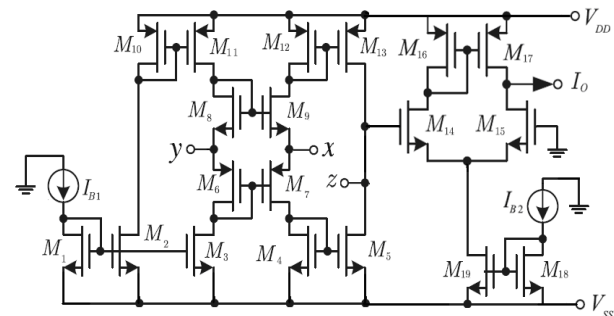


**Fig.1: (a) Symbol (b) Equivalent circuit**

## 2.1 Proposed CMOS Model of CCCCTA

The proposed CCCCTA consists of two principal building blocks: a current controlled second generation current conveyor (CCCII) circuit and an operational transconductance amplifier (OTA) circuit. The proposed realization of the CCCCTA in a CMOS technology to achieve a wide-range of frequency responses is shown in Fig. 2. The circuit implementation consists of the

mixed translinear loop ( $M_6$ – $M_9$ ). The mixed loop is DC biased by using current mirrors ( $M_1$ – $M_3$  and  $M_{10}$ – $M_{11}$ ). The output Z-terminal that generates the current from the X-terminal is realized using transistors ( $M_4$ – $M_5$  and  $M_{12}$ – $M_{13}$ ). The simplified transconductance amplifier is realized using transistors ( $M_{14}$ – $M_{17}$ ). The transconductance amplifier is DC biased by using current mirror ( $M_{18}$ – $M_{19}$ ) and transconductance gain can be adjusted by  $I_{B2}$ .



**Fig.2: Proposed CMOS CCCCTA**

## 2.2 CMOS CCCCTA Performance Results

PSPICE simulation on the CMOS implemented circuit of CCCCTA employed the PMOS and NMOS transistors, using the parameters of a 0.35 $\mu$ m TSMC CMOS technology with  $\pm 1.5$ V supply voltages. The aspect ratios of PMOS and NMOS transistors are listed in Table 1.

**Table 1: Dimensions of the MOS transistors**

CMOS Transistors	W( $\mu$ m)/L ( $\mu$ m)
M1-M5	5/0.5
M6-M7	4/0.5
M8-M9	2/0.5
M10-M13	15/0.5
M14-M15 M18-M19	15/1.5
M16-M17	30/1.5

## D.C. ANALYSIS

Fig.4 displays DC characteristics of the proposed CCCCTA, when  $I_{B1} = I_{B2} = 30 \mu$ A. So it is clearly seen that it is linear in  $-500\text{mV} < V_y < 500\text{mV}$ .

Fig.5 and Fig.6 displays DC characteristics of the proposed CCCCTA, when  $I_{B1}=I_{B2}=30\mu A$ . So it is clearly seen that it is linear in  $-500\mu A < I_X < 500\mu A$ .

### A.C. ANALYSIS

A.C. input signal is also applied at input terminal and simulated bandwidths of output terminals are shown in Fig.7. The  $-3$  dB cutoff frequencies of the current gains  $I_z/I_x$  and  $I_o/I_x$  are, respectively, located at 835MHz, 198 MHz, when  $I_{B1}=I_{B2}=50\mu A$ . It has been found that the frequency responses at O-terminal is narrower than those at the Z-terminal, this is due to signal transmission from the Z to O-terminal of transconductance amplifier.

### TRANSIENT RESPONSE

PSPICE simulation is also carried out for sinusoidal inputs of 100KHZ frequency. These results also give a good agreement between expected and experimental results. These results also verify the basic Equation of CCCTA in time domain in Fig. 8. The THD was found low up to the high frequency.

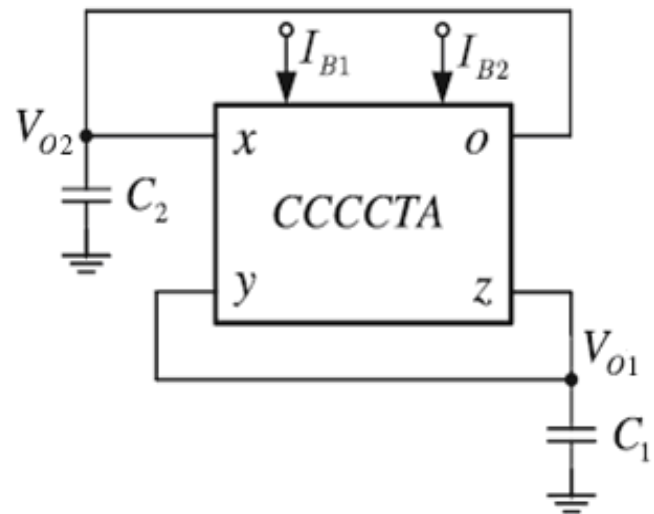
### 3 VOLTAGE-MODE OSCILLATOR

Numerous oscillator circuits using different types of current conveyors and passive components have been reported. Whereas most of the works reported were based on voltage mode outputs. Sinusoidal oscillators play an important role in instrumentation, communication and signal processing applications. Several of the already reported circuits exhibit the Quadrature outputs (with  $90^\circ$  phase shift, whereas some circuits provide multiphase outputs.

#### Circuit Description

The voltage-mode oscillator based on CCCCTA is shown in Fig.3. It consists of a

single CCCCTA and two grounded capacitors.



**Fig. 3: VM oscillator based on the CCCCTA**

### Design and Verifications

#### *Frequency response*

The voltage-mode oscillator of Fig.3 is verified using the model of PMOS and NMOS transistors. The parameters of a  $0.35\mu m$  TSMC CMOS technology with  $\pm 1.5V$  supply voltages have been used. To prove the performance of the voltage-mode oscillator, the PSPICE simulation program used.  $C1 = C2 = 0.01nF$ ,  $I_{B1} = 10\mu A$ , and  $I_{B2} = 330\mu A$  are chosen to obtain the pole frequency of 3.77MHz. The waveforms and fourier plots are shown in Fig.9. The T.H.D. of the proposed circuit at output is within 3%. The frequency of oscillation is found to vary from 0.6MHz at  $I_{B2}=10\mu A$  to 4.25MHz at  $I_{B2}=500\mu A$ , which shows a wide variation in frequency for a variation in  $I_{B2}$  and frequency of oscillation can be independently controlled with the help of  $I_{B2}$ , without any change in condition of oscillation.

# **4 FIGURES**

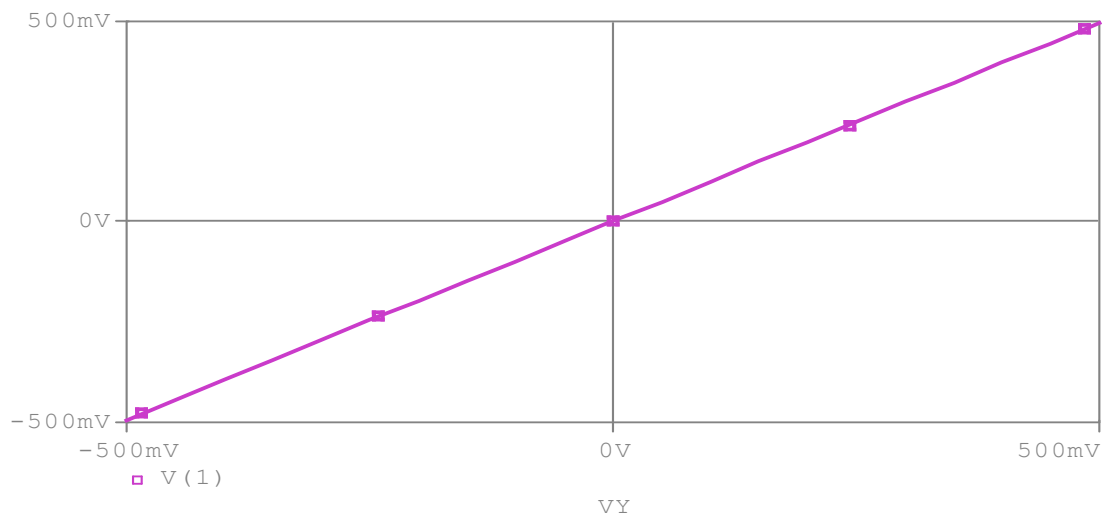


Fig. 4:  $V_x$  vs  $V_y$  plot for CCCCTA

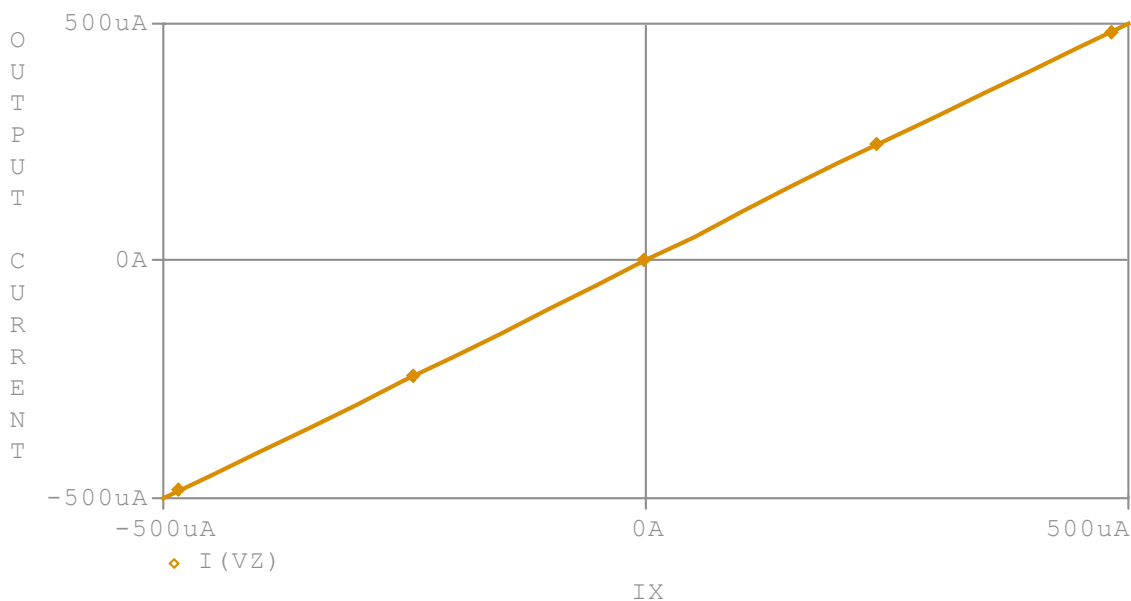


Fig. 5: Output current variation with  $I_x$

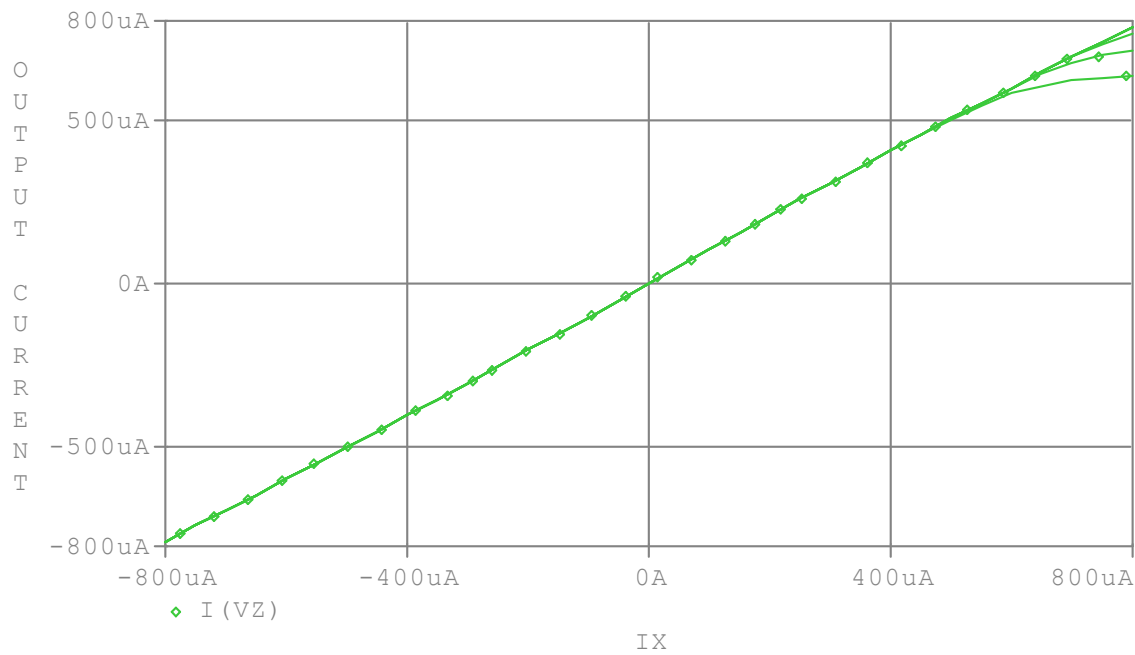


Fig.6: Output current variation with  $I_x$  showing  $I_{B1}$  independence

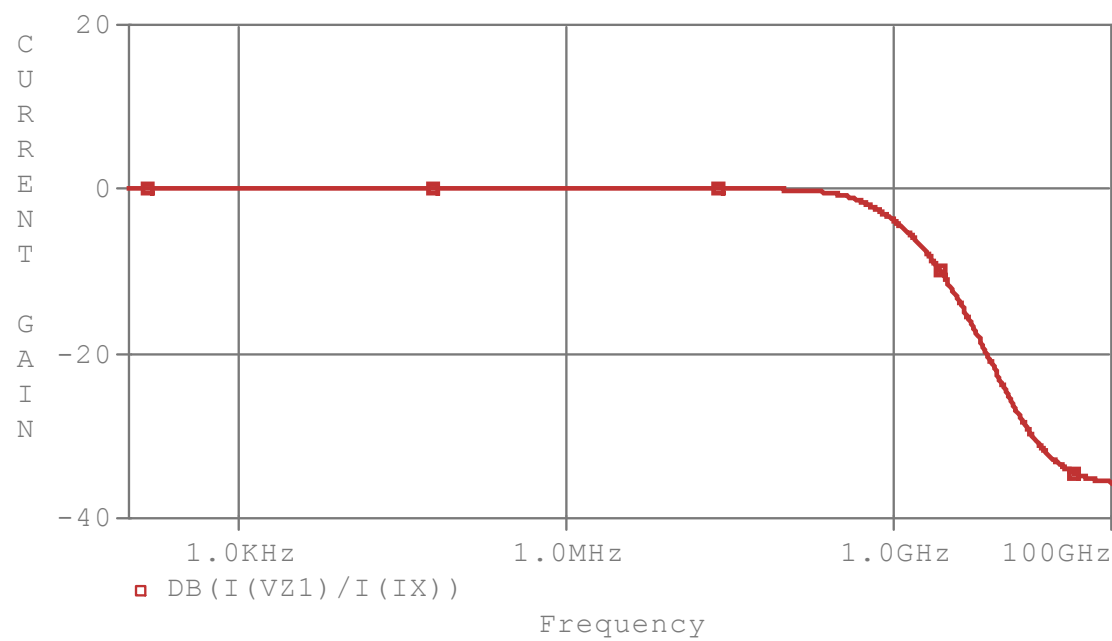


Fig. 7: Current Gain in dB



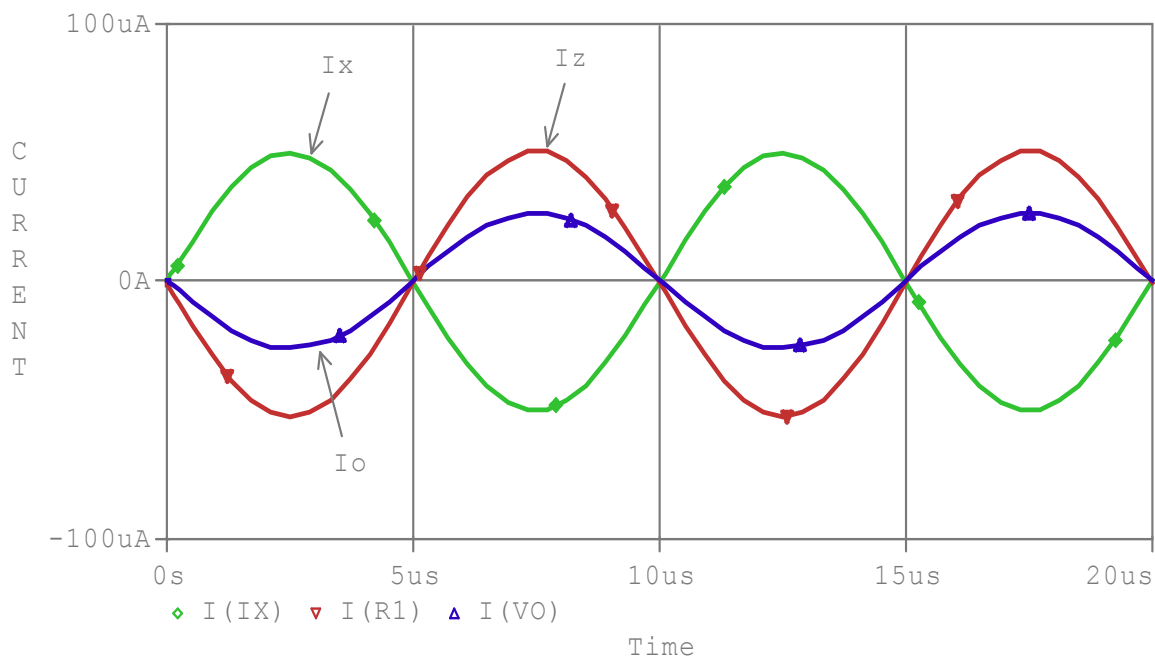


Fig. 8 Input/Output Current responses

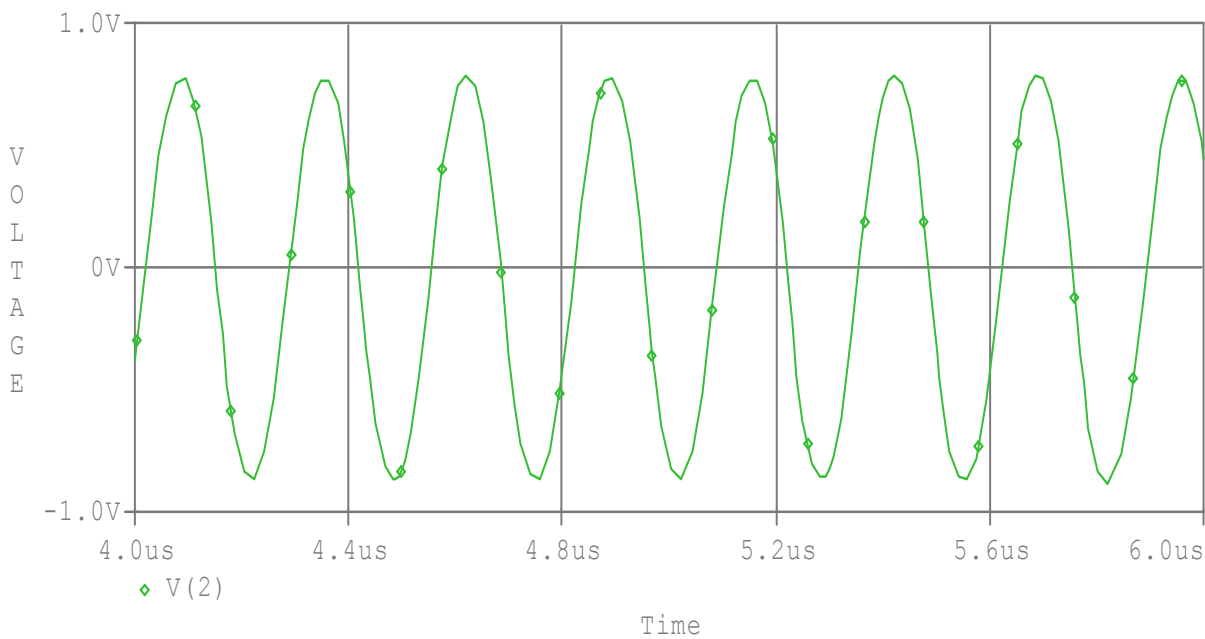


Fig.9: Voltage output ( $V_{O1}$ ) of voltage-mode oscillator

## 5 CONCLUSIONS

Current-mode circuits are undoubtedly the most widely accepted operational devices in continuous time and current mode signal processing. In addition a number of novel circuit functions and topologies have been explored on a front of current mode analogue circuits, opening up wider area of interest. The new building block, called as CCCCTA, has been introduced via this paper. The usability has been proven by the simulation and application examples. They require few numbers of components while electronic controllability is still available, which differs from the recently proposed elements. This novel element is very appropriate to realize in a commercially-purposed integrated circuit. Our future work is to find more applications of the CCCCTA, emphasizing on current-mode signal processing circuits such as multiplier/divider, rectifier, etc.

## 6 REFERENCES

- [1] K. Smith and A. Sedra. (1968) 'The Current Conveyor-a new circuit building block,' IEEE Proc., vol. 56, pp. 1368-1369.
- [2] A. Sedra, K. C. Smith. (1970) 'A second generation current conveyor and its application', IEE Trans. Circuit Theory, pp. 132-134.
- [3] Rajput, S. S., and Jamuar, S. S. (2004). "Low Voltage, High Performance Current Conveyors for Low Voltage Analog And Mixed Mode Signal Processing applications", Analog Integrated Circuits and Signal Processing 41(1), 21-34.
- [4] D. Birolek(2003) 'CDTA Building block for current-mode analog signal processing,' Proceedings of the European conference on circuit theory and design 2003 - ECCTD'03, Krakow, Poland, pp. 397-400.
- [5] S. Maheshwari, I.A. Khan (December 2004) 'Current-controlled current differencing buffered amplifier: Implementation and applications', Active passive electronic components, No.4, pp. 219-227.
- [6] A. ÄU. Keskin, Dalibor Birolek, Erhan Hancioglu and Viera Biolkova (2006) 'Current-mode KHN filter employing current differencing transconductance amplifiers,' AEU International Journal of Electronics and Communications, Vol. 60, No. 6, pp.443-446.
- [7] M. Siripruchyanum, W. Jaikla (2007) 'Realization of current controlled current differencing transconductance amplifier (CCCDTA) and its application,' ECTI electronics and communications, Vol. 5, No.1.
- [8] S. Maheshwari, I. A. Khan, and J. Mohan (2006) "Grounded capacitor first-order filters including canonical forms," Journal of Circuits, Systems and Computers, vol. 15, no. 2, pp. 289-300.
- [9] S. Maheshwari and I. A. Khan, "Current controlled third order quadrature oscillator", IEE Proc. Circuits Devices and System, vol. 152, pp. 605-607, 2005.
- [10] S. Maheshwari, "Electronically tunable quadrature oscillator using translinear conveyors and grounded capacitors", Active and Passive Electronic Components (USA), vol. 26, pp. 193-196, 2003.
- [11] A. U. Keskin and D. Birolek, "Current mode quadrature oscillator using Current Differencing Transconductance Amplifier (CDTA)", IEE Proc.-Circuits Devices Systems, vol. 153, 2006.

# Hybrid Technique for Self Tuning PI Controller Parameters in HVDC Systems

A.Srujana

Research Scholar  
JNT University, Hyderabad  
srujanaphd@gmail.com

Dr. S.V.Jayaram Kumar

Professor  
Jawaharlal Nehru Technological University  
Hyderabad

**Abstract**— Nowadays, due to certain advantages, the HVDC systems are commonly used in long distance transmissions. The major drawback associated with HVDC system is that it takes a longer duration to return to its steady state value after the occurrence of a fault. In a HVDC system, when a fault occurs, the current and voltage will deviate from their normal range and PI controllers are used to maintain its current and voltage at the normal steady state value. Controller parameter tuning plays a significant role in maintaining the steady state current and voltage of a HVDC system. Here, we propose a hybrid technique to self tune the PI controller parameters. The proposed hybrid technique utilizes fuzzy logic and neural network to self tune the controller parameters. The fuzzy rules are generated using different combinations of current error, rate and combined gain. To train the neural network, different combinations of fuzzy gain, proportional gain and integral gain are used. The neural network is trained using a back propagation algorithm. By experimentation it is shown that the system that uses this method takes a very short time to return to its normal steady state. The implementation results show that the performance of the proposed hybrid technique is superior to that of both the self tuning techniques.

**Keywords**- fuzzy logic; HVDC; neural network; fuzzy rules; proportional and integral gain.

## I. INTRODUCTION

Presently, due to economic, environmental, and political limitations which hinder the erection of large power plants and high voltage lines, increasing the power system capacity is often difficult. Hence, to solve the above issues new solutions are sought. One of the most promising solutions suggests the replacement of conventional HVAC transmission technologies by targeted deployment of HVDC (High Voltage Direct Current) ones [1]. Of late, there has been a significant increase in the HVDC systems that interconnect large power systems offering many technical and economic benefits [2].

HVDC is a proven technology and the features presented by it have made it more alluring than AC transmission for certain applications for example long submarine cable links and interconnection of asynchronous systems [1]. Fixed gains PI controllers are commonly used by HVDC systems [3]. The operating in which a HVDC system can be designed are bipolar mode; mono-polar metallic return and mono-polar ground return modes [5]. Charging the capacitance of a transmission line with alternating voltage is not necessary for

HVDC, so it has the advantage of providing more efficient long distance transmission [21]. System interconnection use of HVDC transmission link has not attracted much awareness [4]. In power transmission systems, HVDC converters have the unique virtues of large capacity and fast controllability [18].

In recent years, because of the development of power electronics, an active role is played by HVDC transmission link based Voltage source converters (VSC), using self-commutated valves (IGBTs, IGCTs and GTOs) in improving the electricity transmission and distribution system [9]. VSC-HVDC system is one of the most modern HVDC technologies, and it incorporates two VSCs, one function as a rectifier and the other as an inverter [8].

In power distribution and transmission systems, line to ground, line to line, double line to ground, and three-phase to ground are the possible faults [11]. The literature presents lot of fault detection techniques. The method based on the sequence components of the fundamental frequency of the post-fault current and voltage is an example for this [14]. A general Fault Detection and Diagnostic scheme consists of two phases, namely symptom generation and diagnosis [1]. So as to accomplish this, by executing modern control strategies the power system must be maintained at the preferred operating level [7]. Contemporary controls which are based on Artificial Neural Network, Fuzzy system and Genetic algorithm are found to be quick, and reliable. Hence, they can be employed for protection against the line faults [13].

Generally, the controller is tuned adaptively to perform the controlling effectively. However, because a single technique is deployed for this purpose, the effectiveness remains a challenge as the necessity and complexity of HVDC system peaks. To overcome this issue, in this paper, we propose a hybrid technique by means of which the PI controller that controls the HVDC system is self tuned whenever a fault occurs. The rest of the paper is organized as follows. Section II reviews the related works briefly and section III details the proposed technique with sufficient mathematical models and illustrations. Section IV discusses implementation results and Section V concludes the paper.

## II. RELATED WORKS

Chi-Hsiung Lin [22] has discussed the difference between two faults in an HVDC link. A misfire fault in the rectifier

valve and inverter valve are the two faults that have been compared. A dynamic simulation analysis has disclosed that the resultant phenomena are not the same. A misfire fault in the rectifier valve creates a substantial torsional torque in a turbine generator adjoining the inverter station whenever the natural torsional modes are disrupted by the power disturbance which it induces on the rectifier side of system frequency. Conversely, a misfire fault in an inverter valve attempts to create commutation breakdown in converters which in turn causes HVDC link failure. HVDC link failure if it happens radically affects the rectifier and inverter sides of the generator.

Vinod Kumar *et al.* [23] have presented a HVDC transmission system which operates with speed and precision in a weak ac system and they have analyzed the control strategy and performance of this system, which has been controlled by employing fuzzy. Under oscillations and huge deviations of the input power, the system has been capable of feeding a weak or even dead network. The competence of the link under a variety of disturbances was optimized with the help of the fuzzy logic-based control of the system. Fundamental building blocks that exist in a typical HVDC system have been made available by the proposed model for use by individual users to build their own models. For synchronizing the firing pulses to the HVDC converter, the DQ-type of phase-locked-loop presented has been a specific contribution of the proposed method. Supplying a clean sinusoidal synchronizing voltage from a contaminated and harmonic imprecise commutation voltage has been made possible by this gate-firing unit. The capability of the proposed fuzzy logic based HVDC system to operate steadily and recover steadily in case of short circuit faults, and its obvious merits have been proved by PSCAD/EMTDC based simulations.

Mohamed Khatir *et al.* [24] have discussed that the relative strength of the AC system which connects a HVDC link considerably affects its functioning. Yet, the relative strength of the AC system compared to the capacity of the DC link has a major effect on the interaction between the AC and DC systems and the problems connected with it. In an HVDC inverter following AC system fault in line commutated thyristor inverter feeding a weak AC system, the effect of the DC control on recovery from AC system fault produced commutation failures has been investigated by the proposed method. The study system has been subjected to the AC system fault known as Single phase ground fault. Using MATLAB Simulink, simulation studies have been performed.

Mohamed Khatir *et al.* [25] have discussed that HVDC converter topology type capacitor commutated converter (CCC) are suitable for utilization in long distance transmission via cables. The proposed method has the potential to be employed in HVDC transmission across large bodies of water. The proposed technology of the Capacitor Commutated Converters (CCC) has been presented and its advantages in high power transmission have been illustrated. By employing PSCAD/EMTDC the transient performance evaluations has been presented. From the primary CIGRE HVDC Benchmark model the system has been derived. The results have revealed the enhanced performance of a CCC link in terms of increased

transmission capacity and enhanced stability of the AC network, when it is linked to a very weak AC system.

Bandarabadi *et al.* [9] have discussed the fault-ride through capability improvement possibility through utilization of VSC-HVDC link to transmission network in connection of 160 MW wind farm. 80 individual 2 MW permanent magnet synchronous generators that comprise the 160 MW wind farm has been divided into 4 groups with 40 MW nominal powers. At the time of wind speed fluctuations and after repairing the grid side faults the voltage at the transmission network terminal has to be re-instituted with reduced power losses. Supporting the voltage of transmission network side has also been vital for the VSC-HVDC at the time of short circuit faults in the main grid which is also called as fault ride-through capability improvement. Both uneven speed operations in wind farm network and fault ride-through capability improvement in transmission network have been stressed by the proposed technique. By means of simulation carried out in the PSCAD/EMTDC software, the behavior of the wind farm, transmission voltage and dc voltage for diverse changes in wind speed and three-phase short circuit fault have been studied. The simulation results have proved the performance of the connection method and the improvement in the fault ride-through capability.

Khatir Mohamed *et al.* [26] have presented the steady-state and dynamic performances obtained during step changes of the active and reactive powers, balanced and unbalanced faults in a HVDC transmission system that is based on VSC. It has been shown that fast and satisfactory dynamic responses of the proposed system have been provided by the proposed control strategies in all cases. It has been evident from the simulation, that the VSC-HVDC is capable of performing fast and bi-directional power transfer. It has also been evident that, except for a small fluctuation, the transmitted power can be kept constant at the time of a single-phase fault. Conversely, at the time of a three-phase fault, the power flow by the DC link has been significantly reduced by the voltage at the converter terminals. There has been a quick recovery to usual operation after the fault has been cleared.

Lidong zhang *et al.* [27] have presented a control method of grid-connected voltage-source converters (VSCs). This method has been expected to be of most significance in high-voltage dc (HVDC) applications, though it can be usually applied for all grid-connected VSCs. The proposed method has made use of the internal synchronization mechanism in ac systems, in principle, similar to the operation of a synchronous machine which is different from the preceding control methods. By employing this type of power-synchronization control the instability due to a standard phase-locked loop in a weak ac-system connection has been prevented by the VSC. Furthermore, a VSC terminal has been capable of providing strong voltage support to weak ac system like a normal synchronous machine. By analytical models and time simulations the control method has been proved.

### III. NEURO-FUZZY SELF TUNING PI CONTROLLER IN HVDC

In this paper, for tuning PI controller parameters in HVDC system both normal and abnormal conditions are considered. During normal condition the current remains at its reference

value and when a fault occurs in the system, the current value increases and at that moment the PI controller parameters are tuned and this makes the current to remain at its reference value. Here we use a hybrid technique to tune the PI controller parameters in HVDC. First the error and rate values are calculated from the current value and they are given as input to the fuzzy logic and the fuzzy logic produces a combined gain as the output. The fuzzy gain is given as the input to the neural network which in turn gives the proportional and

integral gain as the output. By using this proportional and integral gain, the controller parameters are adjusted and makes current to remain stable.

#### A. System Model

HVDC system model considered in our method is shown in Figure 1. HVDC systems are commonly used for long distance transmission and its major problem is due to the fault that occurs in the system.

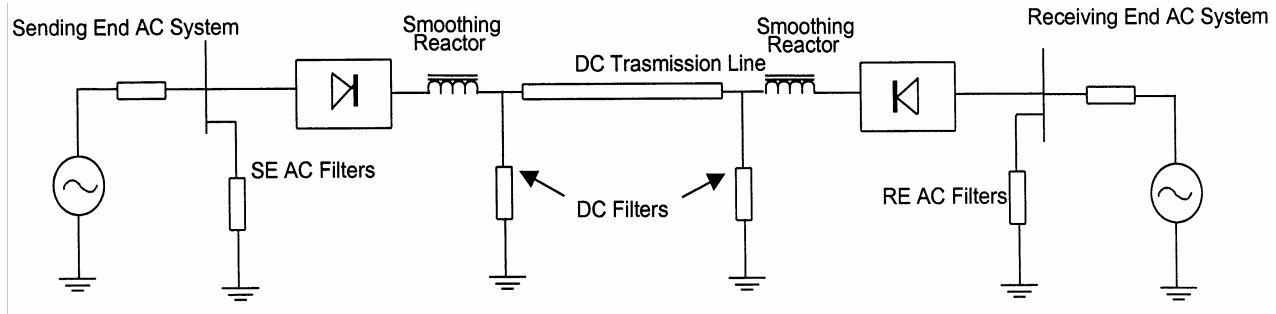


Figure 1. HVDC system model

The faults considered in our system are

- i. Single Line to Ground fault
- ii. Line to Line fault

#### i. Single line to Ground fault

The single line to ground fault is a very common fault in HVDC systems. During this fault the current value gets increased and the corresponding voltage decreases.

#### ii. Line to Line fault

The line to line fault occurs between two transmission lines. This fault is one of the common faults that occur in overhead transmission lines.

When a fault occurs in the system the current value increases and due to this increased current more problems occur in the system. To control this current we used a hybrid technique which is a combination of fuzzy logic and neural network. The fuzzy logic is trained by giving error and rate values are given as its input.

The error and rate values always depend on the current. If the current value is normal then the error is zero and if current increases the error also increases. The error and rate are calculated by using the equations given below.

$$\Delta I_{dc} = I_{ref} - I_m \quad (1)$$

$$\dot{\Delta I}_{dc} = \frac{(\Delta I_{dc} - \Delta I_{pv})}{\Delta T} \quad (2)$$

$$E = G_1 \cdot (\Delta I_{dc}) \quad (3)$$

$$R = G_2 \cdot (\dot{\Delta I}_{dc}) \quad (4)$$

where,  $I_{ref}$  is the reference current,  $I_m$  is the measured current,  $\Delta T$  is the sampling rate,  $\Delta I_{pv}$  is the previous value of error, and  $G_1, G_2$  are the gains for normalization.

By using the formulas the error and rate are calculated and these calculated values are given as input to the fuzzy logic.

#### B. Obtaining Fuzzy Gain

The fuzzy logic is used here, to obtain the combined gain. The current error and rate are the inputs given to the fuzzy logic for rectifier pole controller and inverter controller and its output is the combined gain.

The error and rate are given as input to the fuzzy logic and the combined gain is obtained as its output. For obtaining this fuzzy logic, the generation of fuzzy rules and training are important processes and these processes are explained in section III.E and III.F respectively. By giving any value of error and rate as input to the fuzzy the related combined gain can be got as its output. When the current value changes the combined gain also changes accordingly. Then, the fuzzy output is given as an input to the neural network and the proportional and integral gain are obtained from the neural network as outputs. Based on the change in the combined gain that is given as input to the neural network, the proportional and integral gain values will change.

#### C. Obtaining PI Controller Parameters from Neural Network

Artificial neural networks (ANNs) are excellent tools for complex manufacturing processes that have many variables and complex interactions. Basically, neural network consists of three layers, namely input layer, hidden layer and output layer. In our model, input layer has one variable, hidden layer has n variables and output layer has two variables.

The configuration of the network used is shown in Figure 2.

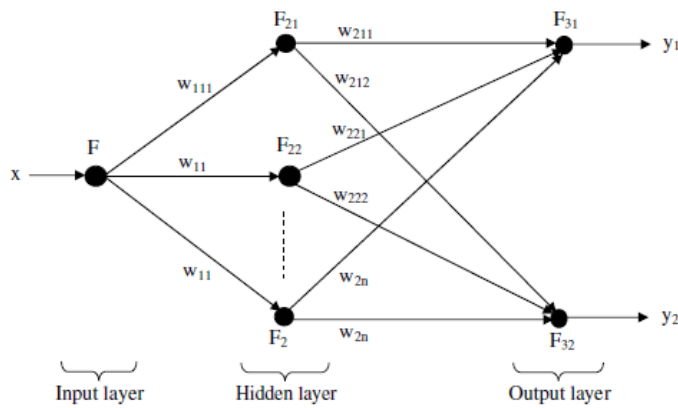


Figure 2. Single input and two output neural networks to obtain proportional and integral gain

The input to the neural network is the fuzzy gain and its outputs are proportional and integral gain. For operation, the neural network must be trained using a training data set. The training of neural network is explained in section III.F. Once the network is trained a general model is created for the relationship between its input and output. So, when the combined gain value is given as an input to the network the related proportional and integral gain will be got as the output.

By using this technique, the PI controller parameters are tuned automatically and the current is made to remain stable, even if any fault occurs in the system it returns to its stable value in a short time.

#### D. Fault Clearance

During normal condition, current will remain in its steady state value and so when the function checks the value, it decides that no change is necessary in the system. When a fault occurs in the system the current will increase suddenly. This time when the function checks the current value, it identifies the increase in current and calls our technique. By using our technique, the error and rate values are calculated from the current value and they are given as input to the fuzzy logic. By giving error and rate values as an input to the fuzzy logic we get fuzzy gain as the output. This gain value is given as input to the neural network and the neural network gives proportional and integral gain as output. By using the proportional and integral gain from the neural network the current values are calculated using the equation given below.

$$I_{out} = K_p e + K_I \int_0^T e dt \quad (5)$$

where,  $I_{out}$  is the output of the PI controller. After calculating the current values the function checks whether the steady state value of current is reached or not. If the current has reached the steady state value then the function stops its process. If the current has not reached the steady state value then the function will repeat the process by calculating the error and rate values again and give them as input to the fuzzy. This process continues until the current reaches the steady state values i.e., until error value reaches zero.

In our method the fault that occurs in both rectifier and inverter sides of the HVDC system are considered. In inverter side the maximum fault current for line to ground fault is 2 KA and in line to line fault maximum fault current value is 2.5KA. In rectifier side the maximum fault current value is 1.5 KA for both single line to ground fault and line to line fault. When a fault occurs in the system the current reaches its maximum value and voltage becomes value. For maintaining the current at its normal value the current value must be reduced and voltage value must be increased. By using our technique  $K_p$  and  $K_i$  values are adjusted, to make the current reach its normal value. By using this method, when a fault occurs in the system the current can be made to return to its normal value within a fraction of a second.

#### E. Generation of Fuzzy Rules

For training the fuzzy logic, training data set and fuzzy rules are generated. The faults which are considered for training fuzzy logic are line to line fault and line to ground fault in both rectifier and inverter side. By considering these faults the input variables are selected and based on that the training data set for fuzzy logic is generated. Inputs are fuzzified in to three sets i.e.; large, medium and small and outputs are very large, large, medium, small and very small. The membership grades are taken as triangular and symmetrical. Fuzzy rules are generated by considering both normal and abnormal conditions.

For different combination of input variables the generated fuzzy rules are shown in table I. After generating fuzzy rules the next step is to train the fuzzy logic.

TABLE 1. FUZZY RULES

Sl.no	Fuzzy rules
1	if, E=large and R=large, then G=very large
2	if, E=large and R=medium, then G=large
3	if, E=large and R=small, then G=small
4	if, E=medium and R=large, then G=large
5	if, E=medium and R=medium, then G=medium
6	if, E=medium and R=small, then G=large
7	if, E=small and R=large, then G=small
8	if, E=small and R=medium, then G=large
9	if, E=small and R=small, then G=very small

For training fuzzy logic the first step is to generate the training data set. This training data set is generated by calculating error and rate values for different current values. To perform the process, a current dataset  $I$  is generated within the current limit  $[I_{max}, I_{min}]$ . The elements of current dataset are given by  $I = \{I_{min}, I_{min} + I_T, I_{min} + 2I_T, \dots, I_{max}\}$  where,  $I_T$  is a threshold to generate elements in a periodic interval. For every current value the error and rate values are calculated. By using the calculated values fuzzy data set is generated. By using the generated data set the fuzzy logic training process is performed.



#### F. Neural Network Training

The first process for training the neural network is the generation of training data set. Training dataset is generated for training neural network with different combinations of fuzzy gain, proportional and integral gain. For generating training dataset set of fuzzy gain, proportional gain and integral gain are selected and this dataset is used for training the neural network. After generating the training data set, the network is trained using a back propagation algorithm.

The neural network is trained using generated data set. For training neural network back propagation algorithm is used and steps for training neural network are explained below in detail.

The training steps are given as follows:

**Step 1:** Initialize the input weight of each neuron.

**Step 2:** Apply a training sample to the network.

**Step 3:** Determine the output at the output layer of the network.

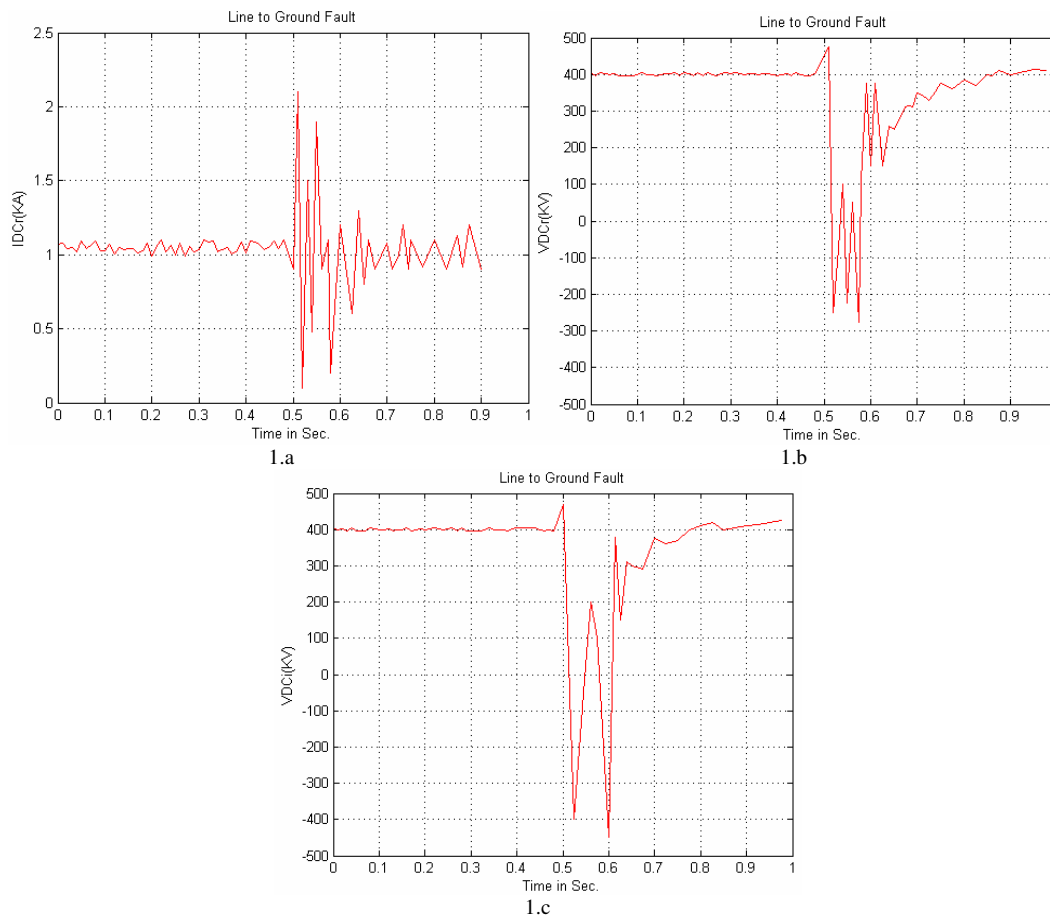
**Step 4:** Determine the value of  $K_p$  and  $K_i$  using the actual output of the network.

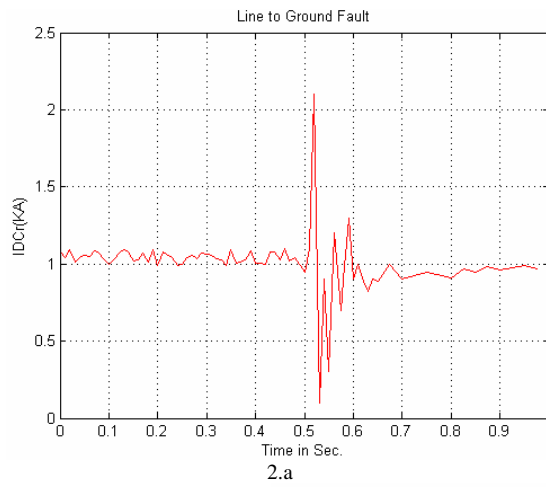
**Step 5:** Repeat the iteration process till the output reaches its least value.

Once the training is completed, the network is ready to tune the control parameters of PI controller and when fuzzy gain changes the network output also changes to maintain the current within the normal range.

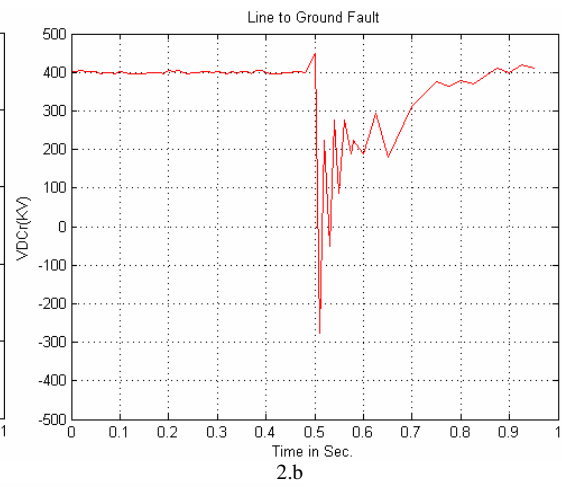
#### IV. RESULTS AND DISCUSSION

The proposed technique was implemented in the working platform of MATLAB 7.10 and its operation was simulated.

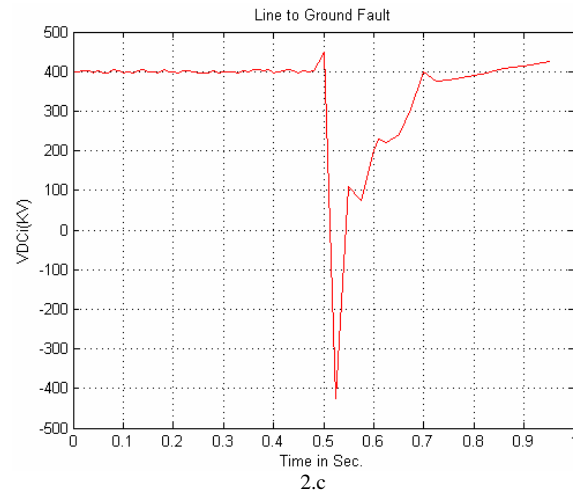




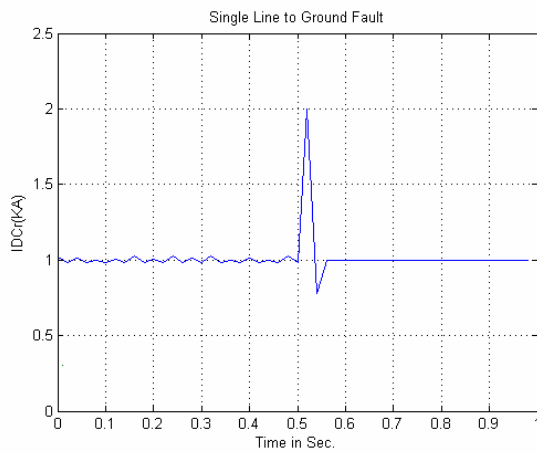
2.a



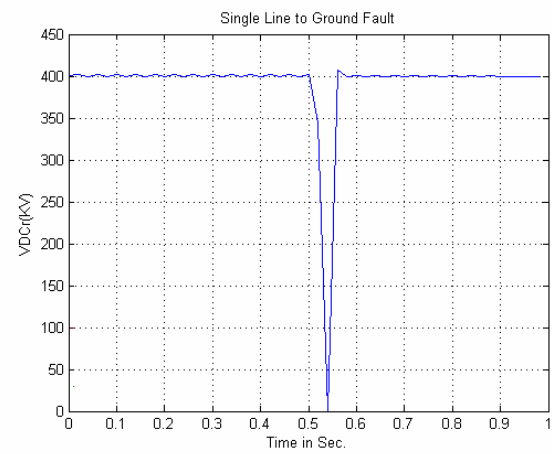
2.b



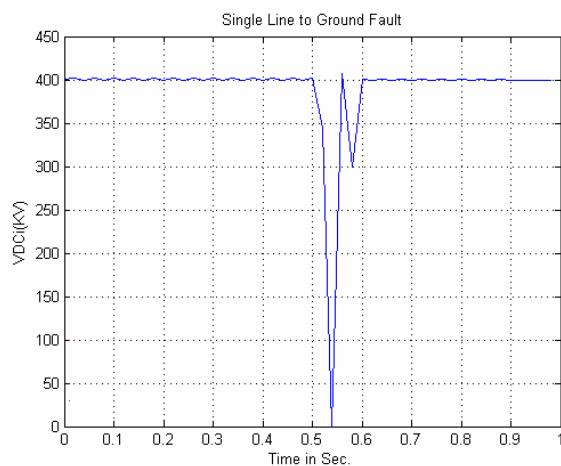
2.c



3.a

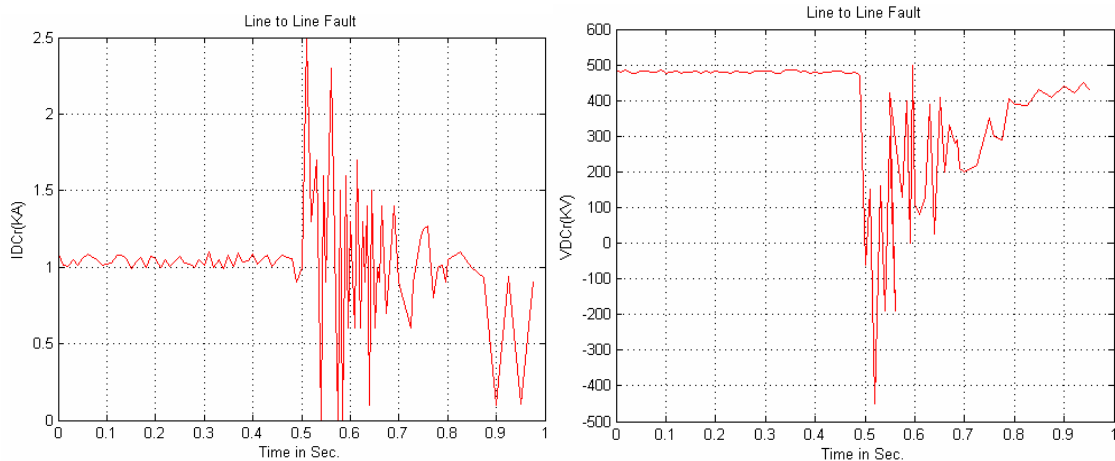


3.b



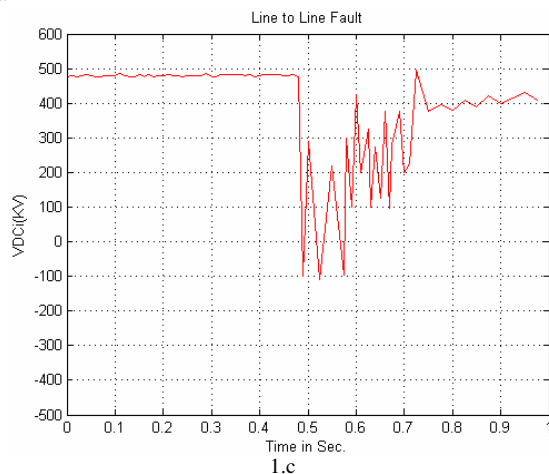
3.c

Figure 3. Performance comparison between (1) conventional, (2) the fuzzy-based and (3) the hybrid PI controller self tuning technique in clearing single line to ground fault at inverter.

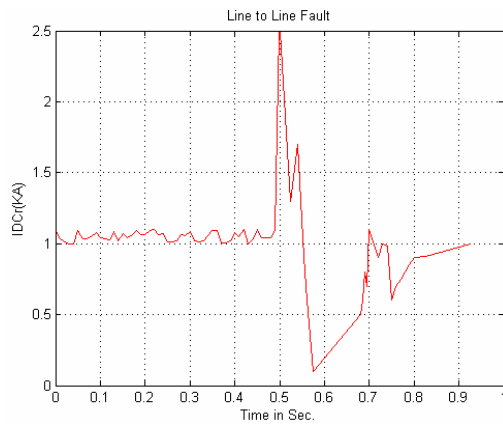


1.a

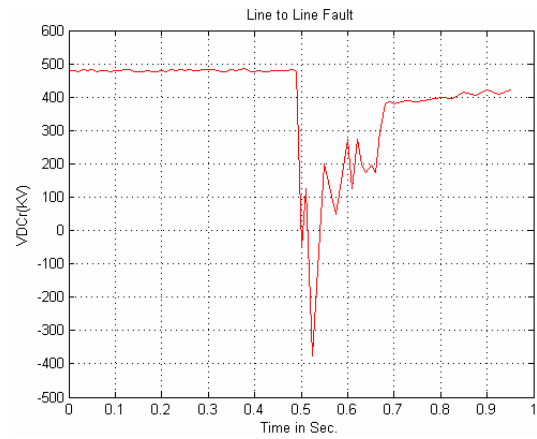
1.b



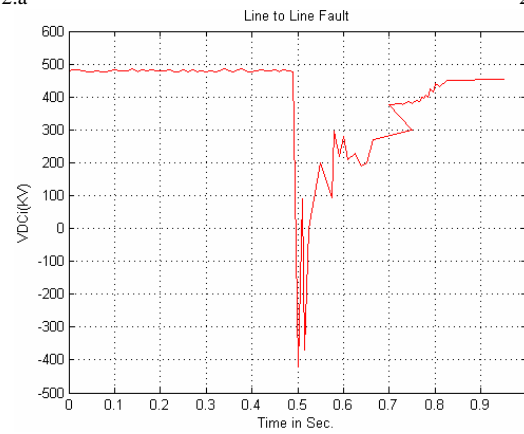
1.c



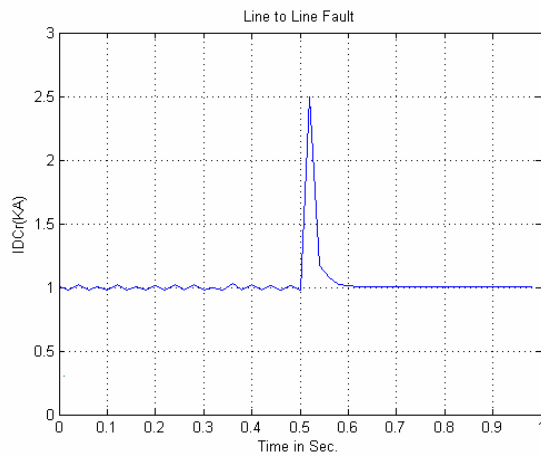
2.a



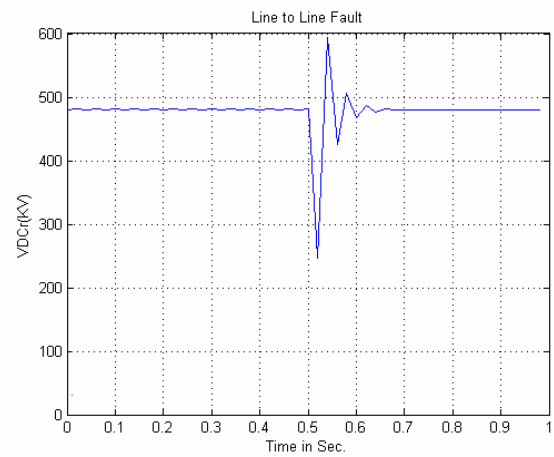
2.b



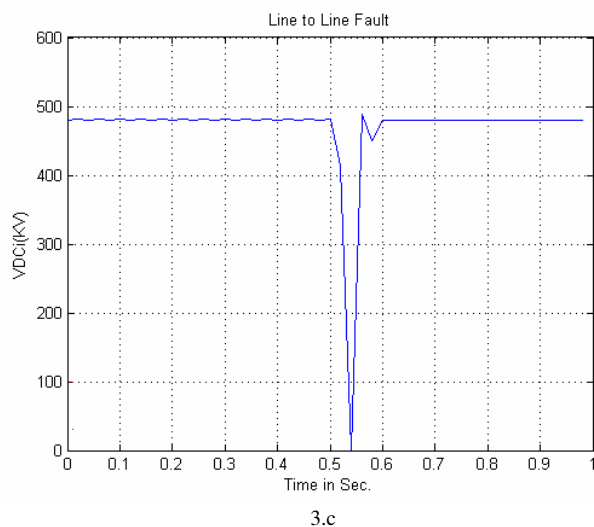
2.c



3.a

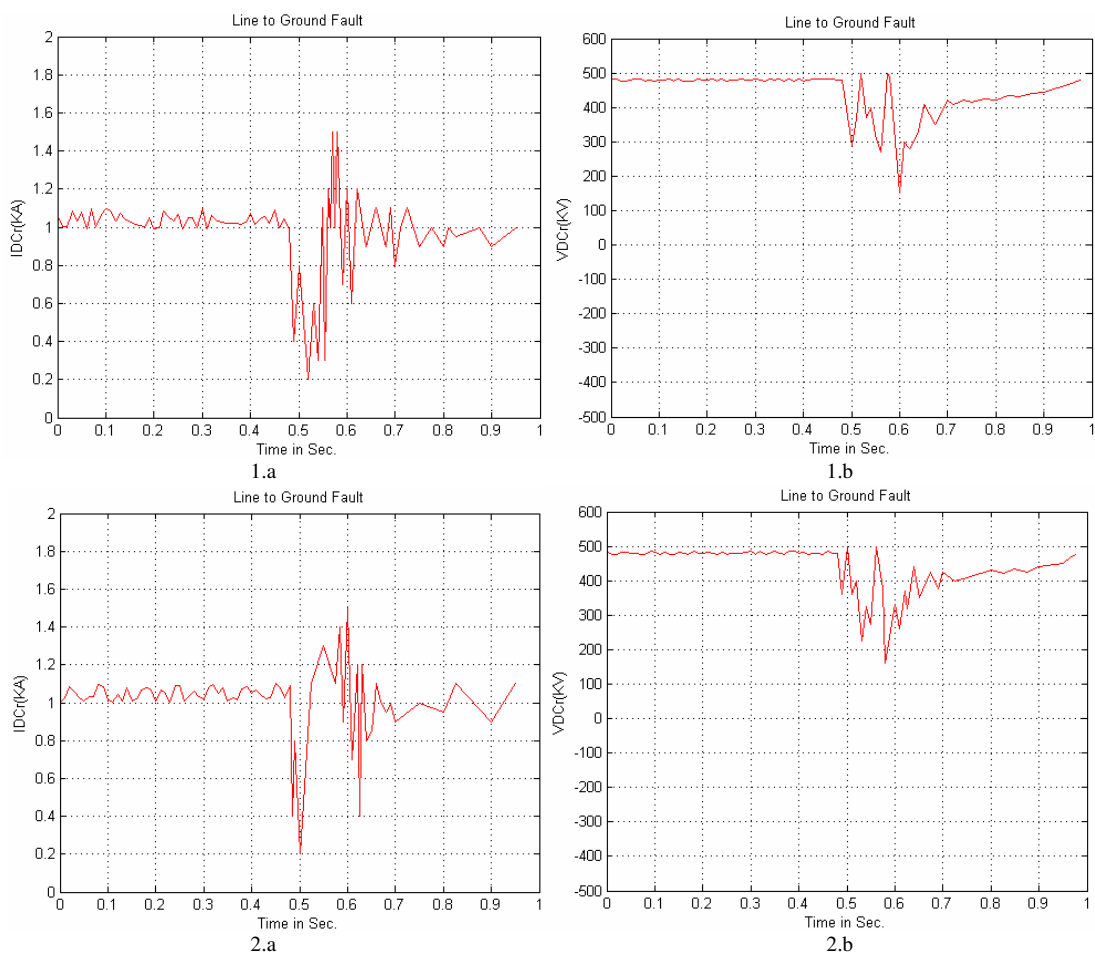


3.b



3.c

Figure 4. Performance comparison between (1) conventional, (2) the fuzzy-based and (3) the hybrid PI controller self tuning technique in clearing line-to-line fault at inverter.



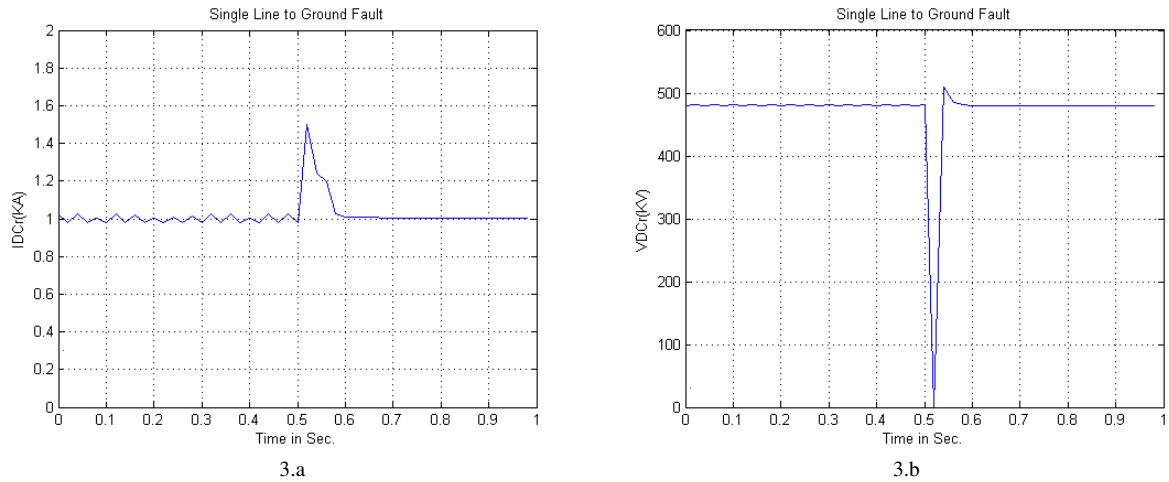
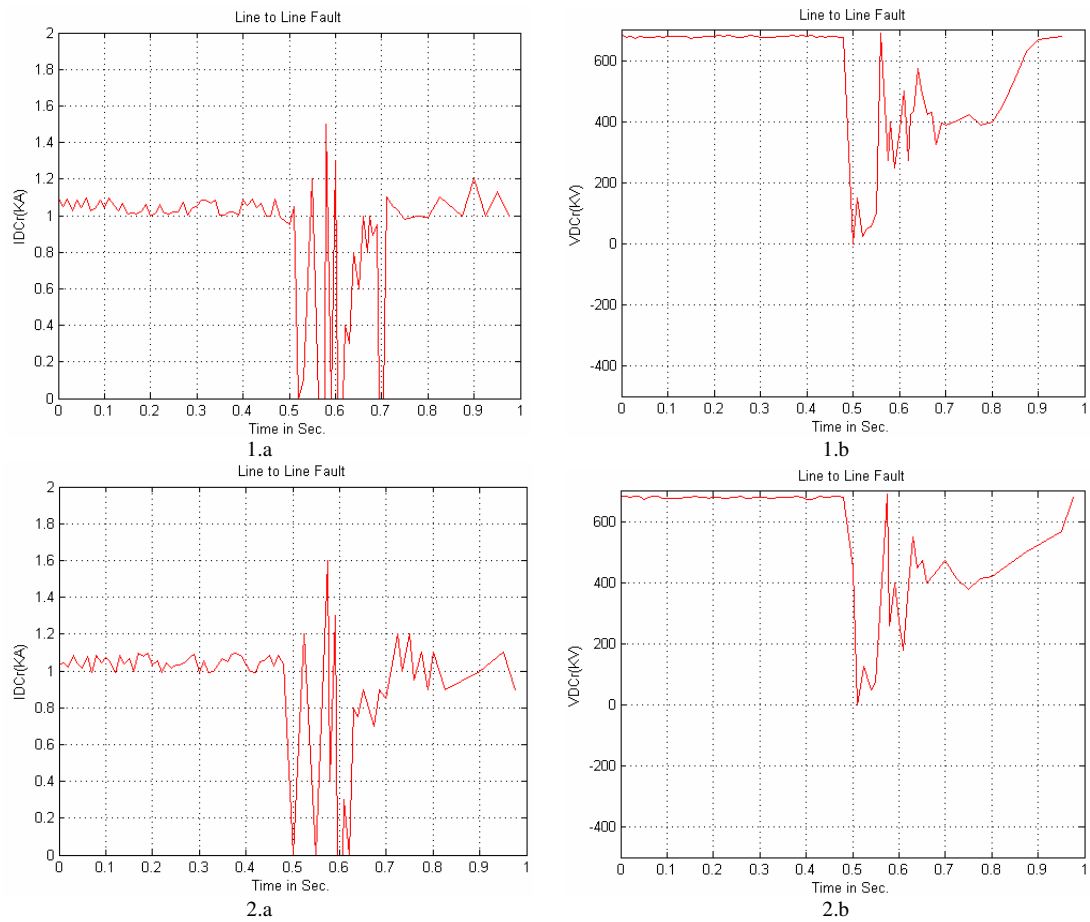


Figure 5. Performance comparison between (1) conventional, (2) the fuzzy-based and (3) the hybrid PI controller self tuning technique in clearing single line-to-ground fault at rectifier.





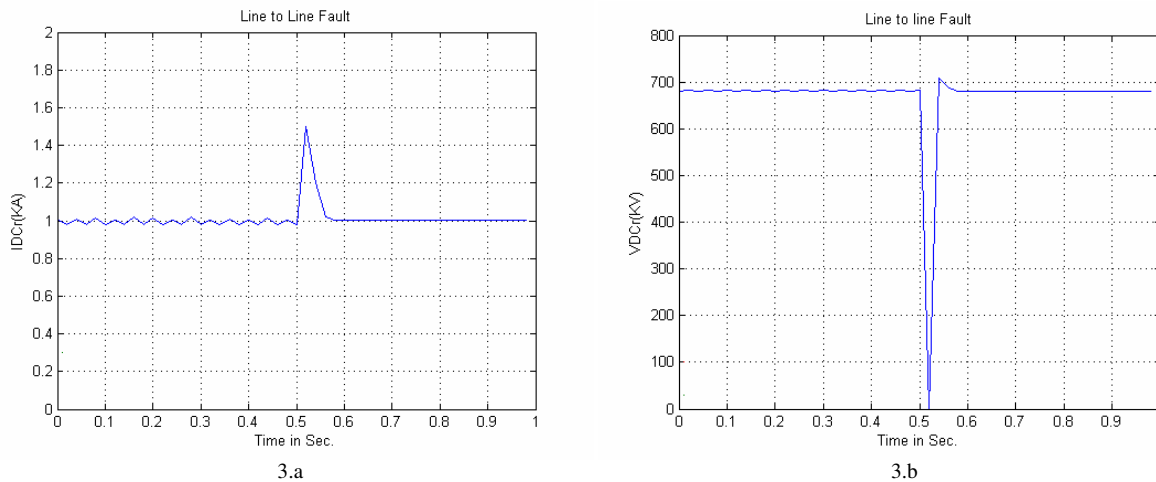


Figure 6. Performance comparison between (1) conventional, (2) the fuzzy-based and (3) the hybrid PI controller self tuning technique in clearing dc line-to-line fault at rectifier.

Only the technique was implemented by MATLAB coding and the model and its operation were considered from [28]. The performance of the proposed technique was compared with the conventional self tuning technique and fuzzy-based self tuning technique. From the results, it is evident that the proposed technique takes considerably less time to stabilize the system than the other existing techniques with which it was compared.

## V. CONCLUSION

In this paper, a neuro-fuzzy hybrid technique to self tune the parameters of the PI controller in a HVDC system, was proposed. Faults which are considered in our system are line to line fault and line to ground fault of both rectifier and inverter sides. When a fault occurs in the system the current and voltage increases and by using this neuro-fuzzy hybrid technique, the system voltage and current can be made to return to their stable values within a fraction of a second. The performance of the system was evaluated from the implementation results. The implementation results showed that the fault clearance time of the hybrid technique is very low compared to conventional methods and fuzzy based self tuning methods. Thus it was proved the proposed technique makes the controlling of HVDC systems significantly more effective than other conventional self tuning techniques.

## REFERENCES

- [1] A. L'Abbate and G. Fulli, "Modeling and Application of VSC-HVDC in the European Transmission System", *International Journal of Innovations in Energy Systems and Power*, Vol.5, No.1, pp. 8-16, April 2010.
- [2] N.M.Tabatabaei and N.Taheri, "Damping Function of Back to Back HVDC Based Voltage Source Converter", *International Journal on Technical and Physical Problems of Engineering*, Vol.1, No.2, P.p.1-7, 2010.
- [3] Narendra Bawane, Anil G. Kothari and Dwarkadas P Kothari, "ANFIS Based HVDC control and Fault Identification of HVDC converter", *HAIT Journal of Science and Engineering*, Vol.2, No.5-6, pp.673-689, 2005
- [4] H.D. Mathur and H.V. Manjunath, "Study of Dynamic Performance of Thermal Units with Asynchronous Tie lines using Fuzzy Based Controller", *J. Electrical Systems*, Vol.3, No.3, pp.124-130, 2007.
- [5] A.Manglik, S.K.Verma, R.P.Sasmal, and D.Muralidharan, "Application of Magnetotelluric Technique in selection of Earth Electrode Sites for HVDC Transmission Systems: An example from NE India", *Journal Earth Science India*, Vol.2, No.4, pp.249 - 257, Oct.2009.
- [6] C.Srinivasa Rao, Z.Naghizadeh and S.Mahdavi, "Improvement of dynamic performance of hydrothermal system under open market scenario using asynchronous tie-lines", *World Journal of Modeling and Simulation*, Vol.4, No.2, pp.153-160, 2008.
- [7] S.Ramesh and A.Krishnan, "Fuzzy Rule Based Load Frequency Control in a Parallel AC - DC Interconnected Power Systems through HVDC Link", *International Journal of Computer Applications*, Vol.1, No.4, pp. 62-69, 2010.
- [8] Mohamed Khatir, Sid Ahmed Zidi, Samir Hadjeri, Mohammed Karim Fellah, "Dynamic Performance Of a Back-To-Back HVDC Station Based on Voltage Source Converters", *Journal of Electrical Engineering*, Vol.61, No.1, pp.29-36, 2010.
- [9] Hanif. Livani, Mohsen. Bandarabadi, Yosef. Alinejad, Saeed. Lesan and Hossein. Karimi-Davijani, "Improvement of Fault Ride-Through Capability in Wind Farms Using VSC-HVDC", *European Journal of Scientific Research*, Vol.28, No.3, pp.328-337, 2009
- [10] Uma Vani M, "Damping effects of Supplementary Control Signals for Enhancement of Transient Stability in AC-DC Power Systems", *International Journal of Engineering Science and Technology*, Vol.2, No.7, pp.084-3092, 2010.
- [11] Rasli A Ghani, Azah Mohamed and Hussain Shareef, "ANFIS Approach for Locating Precise Fault Points with Co-ordinated Geometries in a Test Distribution System", *European Journal of Scientific Research*, Vol.35, No.3, pp.461-473, 2009.
- [12] Alberto Borghetti, Mauro Bosetti, Mauro Disilvestro, Carlo Alberto Nucci, Mario Pauloni, Lorenzo Peretto, Elisa Scalla and Roberto Tinnarelli, "Assessment of Fault Location In Power Distribution Networks", *Electrical Power Quality and Utilization Journal*, Vol.8, No.1, pp. 33-41, 2007.
- [13] Ioana Fagarasan and S.St.Iliescu, "Applications of Fault Detection Methods to Industrial Processes", *WSEAS Transactions on Systems*, Vol.7, No.6, pp.812-821, June 2008
- [14] Pan Zhencun, Wang Chengshan, Cong Wei and Zhang Fan, "Single phase-to-ground fault line identification and section location method for non-effectively grounded distribution systems based on signal injection", *Transaction of Tianjin University*, Vol.14, No.2, pp.92-96, 2008.
- [15] C.H.K. Chui, Z. Litifu and B. Kermanshahi, "Value-based HVDC Enhancement Alternative Selection Using System Well-being Analysis", *Iranian Journal of Science & Technology, Transaction B, Engineering*, Vol.32, No.B3, pp 223-234, 2008.
- [16] Nita R. Patne and Krishna L. Thakre, "Factor Affecting Characteristic of Voltage Sag Due to Fault in the Power System", *Serbian Journal of Electrical Engineering*, Vol. 5, No. 1, pp.171-182, May 2008.
- [17] Arthit Sode-Yome, "System and Network Performance Indicators for the Electricity Generating Authority of Thailand: Current and Future Ones", Vol.1, No.1, pp.8-20, 2009.

- [18] S.Todd, A.R.Wood and P.S.Bodger, "An s-domain model of an HVDC converter", IEEE Transactions on Power Delivery, Vol. 12, No. 4, pp.1723-1729, Oct. 1997.
  - [19] S.Ganapathy and S.Velusami, "Design of MOEA based Decentralized Load-Frequency Controllers for Interconnected Power Systems with AC-DC Parallel Tie-lines", International Journal of Recent Trends in Engineering, Vol. 2, No. 5, pp.357-361, Nov 2009.
  - [20] S. B. Warkad, Dr. M. K. Khedkar and Dr. G. M. Dhole, "Impact of HVDC Transmission on Optimal Electricity Nodal Prices: A study in India", International Journal of Engineering and Technology, Vol.2, No.1, March 2009.
  - [21] Asplund, "Sustainable energy systems with HVDC transmission", in proceedings of IEEE General Meeting of Power Engineering Society, Vol.2, pp.2299-2303, June 2004.
  - [22] Chi-Hsiung Lin, "Phenomena Caused by a Misfire Fault in an HVDC Converter Valve and the Impact on a Turbine Generator", Journal of Technology, Vol. 23, No. 2, pp. 93-100, 2008.
  - [23] Vinod Kumar, Joshi, Garg and Bansal, "Intelligent Controller Based Improved Fault Ride- through Capability of HVDC System Connected to Weak ac Grid", Journal of Theoretical and Applied Information Technology, Vol.4, No.3, pp.203-211, March 2008
  - [24] Mohamed Khatir, Sid Ahmed Zidi, Samir Hadjeri and Mohammed Karim Fellah, "Analysis Of Recovery from Commutation Failures in an HVDC Inverter Connected to a Weak Receiving ac System", Acta Electrotechnica et Informatica, Vol. 8, No. 1, pp. 44-50, 2008.
  - [25] Mohamed Khatir, Sid-Ahmed Zidi, Mohammed-Karim Fellah, Samir Hadjeri and Rabie Amiri, "Performance Evaluation of an HVDC Link with a Capacitor Commutated Inverter Connected to a Very Weak Receiving AC Network", Journal of Electrical Engineering, Vol. 60, No.4, pp.209-214, 2009.
  - [26] Khatir Mohamed, Zidi Sid Ahmed, Hadjeri Samir, Fellah Mohammed Karim and Amiri Rabie, "Performance Analysis of a Voltage Source Converter (VSC) based HVDC Transmission System under Faulted Conditions", Leonardo Journal of Sciences, Vol.8, No.15, pp.33-46, July-December 2009.
  - [27] Lidong zhang, Lennart Harnefors and Hans-Peter Nee, "Power-Synchronization Control of Grid-Connected Voltage-Source Converters", IEEE Transactions on Power Systems, Vol.25, No.2, pp.809-820, May 2010.
  - [28] Aurobinda Routray, P. K. Dash, and Sanjeev K. Panda, "A Fuzzy Self-Tuning PI Controller for HVDC Links", IEEE Transactions on Power Electronics, Vol. 11, No. 5, p.p. 669-679, 1996.
- A.Srujana** received the B.Tech Degree in Electrical Engineering from Kakatiya University, Warangal, India in 1998. She received her M.Tech Degree in Electrical Engineering from Jawaharlal Nehru Technological University Hyderabad in 2002. Currently she is pursuing Ph.D from the same University under the guidance of Dr S.V.Jayaram Kumar. Her research interests include Power Electronics and HVDC.
- Dr S.V.Jayaram Kumar** received the M.E degree in Electrical Engineering from Andhra University, Vishakapatnam, India in 1979. He received the Ph.D degree in Electrical Engineering from Indian Institute of Technology, Kanpur, in 2000. Currently, he is a Professor at Jawaharlal Nehru Technological University Hyderabad. His research interests include FACTS & Power System Dynamics.

# Use of Computerized Web-Based Information System For Determining Losses in 15-6.6 KV Feeders in Traditional Electrical Network Management: Case Study Goma Distribution Electrical Network

Ezekiel U. Okike  
Department of Computer Science  
University of Ibadan  
Ibadan, Nigeria  
euokike@gmail.com

Bakunzi G. Joseph  
School of Computer Studies,  
Kampala International University  
Kampala, Uganda

**Abstract—** Electrical energy plays very vital role in modern global economy. The aim of this study is to develop a framework for a Web-Based Information System (WIS) tool for computing losses from 15 – 6.6 KV Feeders in Traditional Electrical Network Management (TENM). The study was conducted in Goma District in the Democratic Republic of Congo. Data were collected from 26 key staff of Goma Distribution Electrical Network who responded to the questionnaires and from metered reading documents used in the study. The study implemented a Computerized Web-Based Information System (CWIS) to compute different losses in Goma electrical distribution network. The CWIS computed technical losses in five 15-6.6KV feeders of Goma electrical distribution network. The study revealed that among the five feeders, feeder 1 (Sud feeder) consumes 1,469,172.6 KWH representing 66.3% of the total annual energy loss while others presented lower annual losses. This is an indication that Feeder 1 is overloaded and needed to be resized or on the alternative, the installation of another overhead cable that will take the half of the load in charge.

*Keywords-* Electrical energy; energy distribution; feeder loss; computerized information system

## I. INTRODUCTION

Modern global economy has rapidly increased by means of the electrical energy. Electrical energy has also penetrated each area of human activities in such way that it has become a second virtual life. Furthermore, the availability of electrical energy and the quality of services (QoS) to consumers shapes and aids national development efforts.

Goma electrical network is located in North Kivu Province East of Democratic Republic of Congo. With the volcanic stones covering the whole Goma town, laying underground electrical cables is very difficult. Hence the whole electrical network is made up with overhead aluminum cables. The electrical energy supplied to Goma town is from Ruzizi hydraulic generation power station located in the neighboring town of Bukavu in South-Kivu province.

In Democratic Republic of Congo (DRC), the company in charge of generation, management, transmission, and distribution of electrical energy is the “Societe Nationale d’Electricite” (SNEL). The head office of SNEL is located in the DR. Congo capital town of Kinshasa which is 2000 kilometers away. Therefore the use of a Computerized Web-based information System (CWIS) can allow managers to use the information resources from Goma electrical distribution network through a communication medium.

Since SNEL still uses manual reading of energy metering and manual information system, a CWIS becomes inevitable. Goma electrical network has only one Transmission Substation being coded as 2x 10MW-70/15 KV, five primary feeders (1 feeder of 6.6kv and 4 feeders of 15kv), and fifty five Distribution Substations with code 15/0.4kv-6.6/0.4kv.

This research is aimed at implementing a Computerized Web-Based Information System that can compute losses based on data entry from manual reading of the metering system, as well as compute technical losses (TL) in the 15kv and 6.6kv feeders.

#### A. Statement of the Problem

The use of manual information system does not show the losses incurred at all the stages of the generation, transmission and distribution of the electrical energy in Goma electrical network system. In addition to this major problem, the manual information system management includes poor data storage, slow and difficult retrieval, inconsistency in data, data redundancy, and the high probability of losing information stored in files on shelves due to damage or theft. Hence the need for a computerized information system which overcomes these problems cannot be over emphasized.

#### B. Objectives of the study

The specific objectives of this study are as follows:

- To implement a Computerized Web-Based Information System (CWIS) for four 15-6.6KV feeder losses computation
- To demonstrate the current state of electrical power supply equipment by showing the electrical power losses in the four 15-6.6KV feeder as a result of power assigned to Goma electrical network consumers by the use of the manual information management system.

#### C. Research Questions

The questions to be answered in this study are:

- To which level (in percentage) can a Computerized Web-based Information System (CWIS) be effectively useful for 15-6.6KV feeder losses computation in

Goma distribution electrical network management?

- What is the level (in percentage) of maximum losses in 15kv and 6.6kv feeders of Goma distribution electrical network?

## II. RESEARCH METHOD

### A. The Traditional Approach to Electric Power

For roughly a century, the developed world has delivered electric power using the same basic four-step approach: 1) generate power in large, centralized plants; 2) step up the power to high voltages and transmit it to regional utilities; 3) step down the power to medium voltages to distribute it locally; 4) step down the power a final time to deliver it to customer premises. (Figure 2.1.)

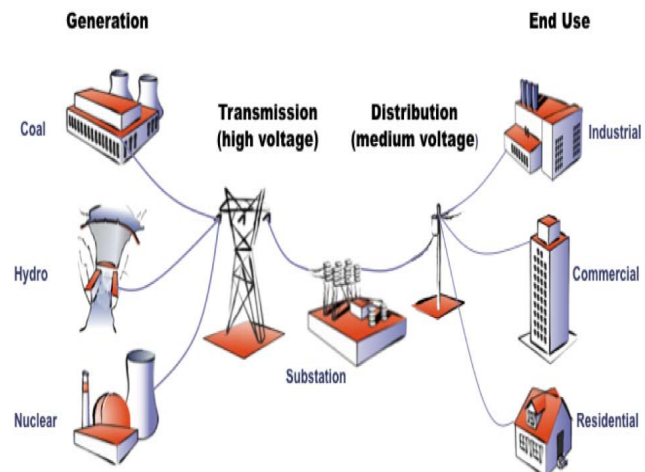


Figure 2.1. Traditional Electric Power Approach

Figure 2.1: The “traditional “ electric power value chain encompassed centralized generation, high voltage transmission, medium-voltage distribution, and end use by industrial, commercial and residential customers (source: Global Environment Fund, 2008)

### B. Goma Electrical Distribution Network Losses Computation procedure and framework

The electrical framework for energy losses in a distribution system is presented in the figure 2.2 .

The steps followed to compute losses at different levels in the framework (figure 2.2) of the electrical distribution network are given below. The primary feeders are five feeders that have a total length of 46 kilometers.

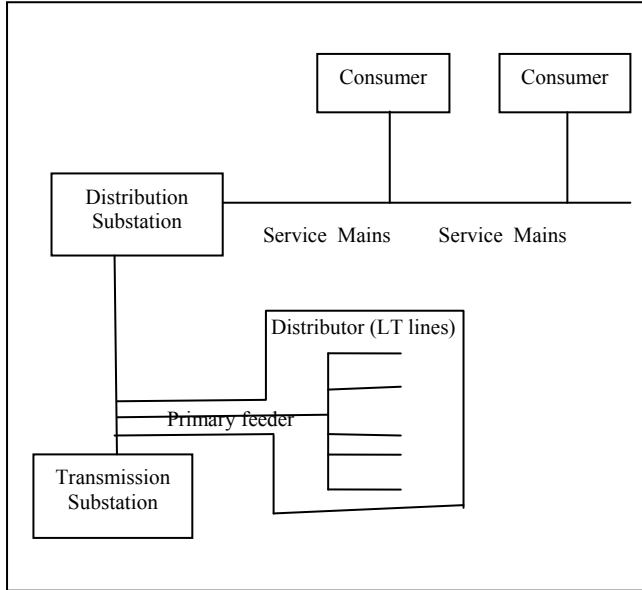


Figure 2.2 Electrical Distribution Network framework

- Utilization Factor (UF) at the existing power factor is given as

$$UF = \frac{\text{Peak load kw}}{\text{kVA capacities of distribution transformer centres}} \quad (1)$$

- Load Factor (LF) and Loss Load Factor (LLF) are given as

$$\text{Peak-Load} = \sqrt{3} V_w (\sum_{n=1}^n I_{pn}) \quad (2)$$

$$\text{Average Load} = \sqrt{3} V_{av} (\sum_{n=1}^n I_{avn}) \quad (3)$$

where  $n$  is the number of feeders,  $V_w$  is working voltage,  $V_{av}$  is average voltage,  $I_{pn}$  is the peak current and  $I_{avn}$  is average current in  $n^{\text{th}}$  feeder. The working voltage  $V_w$  and the peak current  $I_{pn}$  are collected from the metering system.

Hence ,

$$LF = \frac{\text{Average load}}{\text{Peak load}} \quad (4)$$

$$LLF = 0.8 (LF)^2 + 0.2 (LF) \quad (5)$$

- Peak- power loss (PPL)

$$\text{Peak- power loss (PPL) in 11kv line} = 3 I^2 R (UF)^2 \text{ KW} \quad (6)$$

- Annual energy loss

$$\text{Annual energy loss} = \text{PPL} \times 8760 \times \text{LLF KWH} \quad (7)$$

- Average power fed

$$\text{Average power fed} = \sqrt{3} V I (\text{existing pf}) (UF) (LF) \text{ KW with pf as power factor (cosines phi) and KW as kilowatt.} \quad (8)$$

The losses in 15kv lines are calculated as shown below:

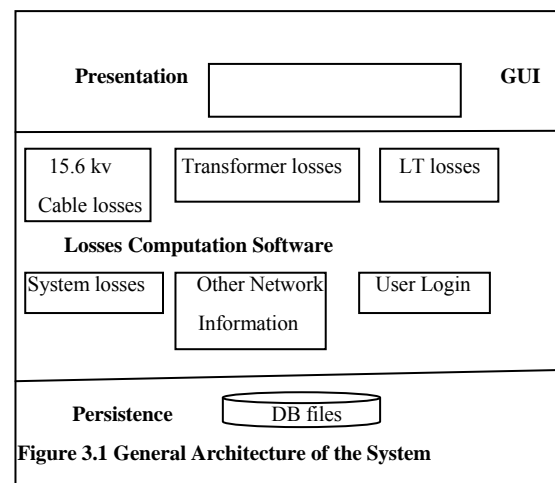
- Percentage 15/6.6 kv line losses

$$\text{Percentage 15kv line losses} = \frac{\text{Power loss in line}}{\text{Average power fed}} \times 100 \quad (9)$$

### III. SOFTWARE IMPLEMENTATION

#### A. General architecture of the System

The general architecture of the system is described by figure 3.1 below:



The Graphic User Interfaces (GUI) were implemented for entering data in the database using PHP 5 with Macromedia dreamweaver 8 programming approaches as proposed by Luke and Laura (2003).

The GUI was designed as form fields for input and display of appropriate data as required by the database for use by the various computational processes. The computational processes were implemented using appropriate international standard empirical formulae (see section 2.2). Forms for reporting errors from the system are generated automatically by the application software. The database was implemented using MySQL. In order to access information in the system, appropriate user authentication and authorization checks were implemented through the system login prompt. The web based capability of the system was implemented using PHP5 with Macromedia dreamweaver8. In all, the entire system requirement are MySQL, PHP5, macromedia dreamweaver, Apache server, and Visio Modeler.

#### IV. SYSTEM RESULT: ELECTRICAL NETWORK DATA PRESENTATION AND INTERPRETATION

##### A. Feeder Losses Presentation and Interpretation

The losses and loads in feeders as results of the system are presented and interpreted in this section. The resistor ( $\Omega$ ), peak load, average load, load factor, loss load factor, capacity transformer power, utilization factor, peak power loss, annual energy loss, cosines phi, and average power fed for each feeder have been computed by the CWIS and presented in table 4.1. The peak power losses (ppl) (kw), the average power (kw), and the 15kv-6.6kv line annual energy loss (kwh) bar charts below show the repartition of loads and losses in the five feeders in order to highlight and guide managers on what decisions for corrective and preventive maintenance may be necessary in order to balance or to reduce losses.

**TABLES 4.1(A-D). FEEDER LINE**

**TABLE 4.1 A**

Feeder Name	Cable code	Length(M)	Ohm (R)	Working volt	Avg volt
Sake	Alu	8761.500	4.505	11.630	10.9
Sous	Alu	16270.000	11.714	5.120	4.8
Nord	Alu	5385.600	3.877	11.630	10.9
Centre	Alu	10758.000	7.745	11.630	10.9
Sud	Alu	5382.300	3.875	11.630	10.9

**TABLE 4.1 B**

Feeder Name	Peak current	Avg current	Peak load	Avg load	Load factor
Sake	98.000	64.000	1974.087	1211.604	0.613
Sous	191.000	140.500	1693.807	1170.528	0.691
Nord	100.000	55.000	2014.375	1041.222	0.516
Centre	70.000	45.000	1410.062	851.909	0.604
Sud	200.000	165.000	4028.750	3123.667	0.775

**TABLE 4.1C**

Feeder Name	Loss Load factor	Capacity Of transformer centres	Utility factor	Peak Power loss	Annual Energy loss
Sake	0.424	5280.000	0.373	18.147	67421.7
Sous	0.520	5725.000	0.295	112.223	511465.2
Nord	0.317	2960.000	0.680	53.874	149664.4
Centre	0.412	6510.000	0.216	5.341	19319.0
Sud	0.635	5350.000	0.753	263.702	1469172.5

**TABLE 4.1 D**

Feeder Name	Total Energy received	Total Energy demded	Cosine phi	Average Power fed
Sake	529.864	345.500	0.837	232.891
Sous	529.864	345.500	0.837	200.472
Nord	529.864	345.500	0.837	306.804
Centre	529.864	345.500	0.837	93.383
Sud	529.864	345.500	0.837	1527.712

From table 4.1 (a-d) the cumulative sum was done by the system to provide the results in table 4.2 (a-b). From table 4.2a, the total length of the 15/6.6 kv cable is shown in the first column and last row as 46.557 kilometers. The next column and last row shows the total peak load computed as 11.121kva (the metered value was 11kva). Other computed values can be seen from the table with the last row of each column accounting for total for total values as follows: average load, utility factor, was equal to 7398.9 kw, utility factor is 2.31991, and total capacity of transformer center is 25.825 kva. However the total power of the Substation which is 20 kva that shows there is extra power of 5.825 kva. The system computed the peak power loss in all the feeders as 453.29 kw, and the annual energy loss as 2,217043.14kwh for the year (2008). The average power fed is computed as 2361.26 kw. The loss load factor and utilization factor have been calculated in order to use their average values



(2.310337divided by 5 and 2.31991divided by 5) in the electrical network losses calculation. It is noted that the average power factor of 0.84 is less than the standard minimum value of 0.90 as stated by Pabla (2005). From these results the study shows that the CWIS can provide information that can be used in preventive and corrective maintenance of the Goma electrical network distribution.

**TABLES 4.2(A-B). SUMMING NUMERICAL VALUES FROM FEEDER READABLE TABLE**

**TABLE 4.2 A**

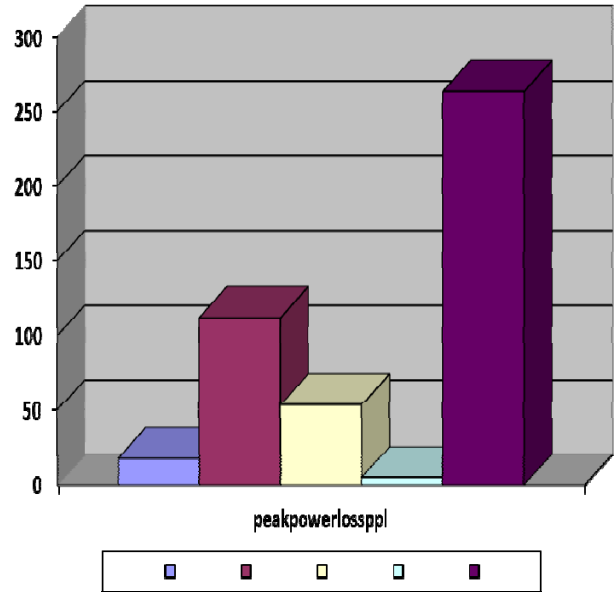
Feeder Name	Length(M)	Peak Load	Average Load	Utility Factor
Centre	8761.5	1974.08	1211.60	0.37
Sake	25031.5	3667.89	2382.13	0.66
Nord	30417.1	5682.26	3423.35	1.35
Sous	41175.1	7092.33	4275.26	1.56
Sud	46557.4	11121.08	7398.93	2.31

**TABLE 4.2 B**

Feeder Name	Loss Load Factor	Capacity Transformer Power	Peak Power loss	Annual Energy Loss	Average power Fed
Centre	0.42	5280	18.14	67421.72	232.89
Sake	0.94	11005	130.37	578886.99	433.36
Nord	1.26	15965	184.24	728551.48	740.16
Sous	1.67	20475	189.58	747870.56	833.55
Sud	2.31	25825	453.29	2217043.14	2361.26

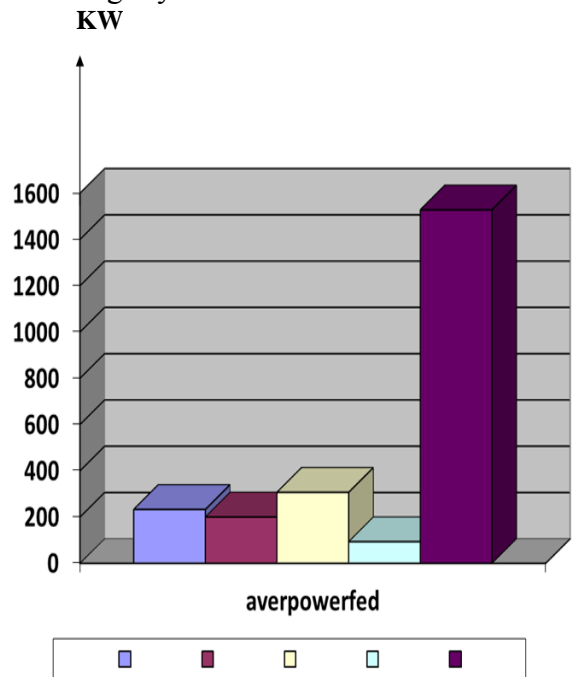
### B. Descriptive Statistics of the System Results

Descriptive statistics of the system results are shown in the bar charts presented in figure 4.1, 4.2, and figure 4.3. The interpretation of results is given below each bar chart.



**Fig 4.1 Sake Sous Nord Centre Sud**

The figure 4.1 reveals that the feeder named “Sud” is most overloaded, and has the highest peak power loss (ppl). At the other hand the feeder named “Center” has the lower load. From this observation the managers can well decide which steps can be taken in order to reduce the power loss in feeder Sud, and how to balance loads on other feeders which are lightly loaded.

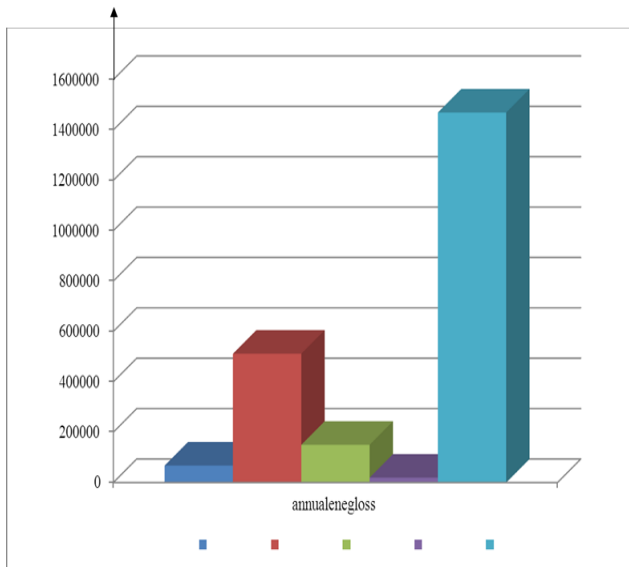


**Fig 4.2 Sake Sous Nord Centre Sud**

The figure 4.2 shows that the same feeder named “Sud” was demanding the highest power compared to others. It reveals that there is a relationship between peak power loss and average power fed. Once again the managers have the precise knowledge of the average power to be cut off from feeder named “Sud”, or the size of the transformer that can supply that average power.

Normally the electrical distribution network manager is interested in how to reduce losses in order to increase profit. The Computerized Web-Based Information System has the capability to compute annual losses to highlight managers about the magnitude of them. The figure 4.3 shows the annual energy losses magnitude for each feeder.

**KWH**



**Fig 4.3 Sake Sous Nord Centre Sud**

The bar charts in figure 4.3 reveal that the same feeder Sud presents the higher annual losses. Out of the total annual energy loss (see table 4.2 - 2217043.14 kwh) the feeder Sud takes itself 1,469,172.6 kwh (see table 4.1) which represents 66.3% of the total annual energy loss. Because the peak current (of 200 ampere) demanded was very high therefore annual energy losses were also increased according to the relation:

Annual Energy losses =  $3 I^2 (UF)^2 \times 8760 \times LLF = K I^2 (UF)^2 LLF$  with K equal to  $3 \times 8760$ .

According to the formular  $LLF = 0.8 (LF)^2 + 0.2$  the Loss Load Factor (LLF) is highly proportional to LF powered by 2. That means that at lower load LLF is small. But at high load LLF is near to one so that the  $I^2$  (the current) is the one highly influencing the annual energy losses. The peak ampere recorded from the metering system being of 200 A, it is visible that the annual energy losses in the feeder named “Sud” must be higher than they are in the other feeders where the peak current range from 70 A to 100 A. The feeder named “Centre” still the one having the lower value of annual losses. The bar charts in figure 5.10 are revealing that there is a correlation between the average power, the peak power loss, and annual energy losses. Managers cannot take decision on how and on which resource to be engaged to reduce losses if they do not know their magnitude. Once again the Computerized Web-Based Information System has shown its capability to compute technical losses (power losses and annual energy loss) for each feeder and has shown how it is a powerful tool to be used in making decisions for preventive and corrective maintenance.

## V. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

### C. Discussion of Findings

Technical losses do occur in electrical feeder distribution networks. Pabla (2005) stated that the typical maximum losses (in percentage) in 15KV and 6.6KV Feeders, should be 4.0%. However, technical losses in the Goma electrical distribution network appear to be far above this range.

Out of the total annual energy loss (see table 4.2 - 2217043.14 kwh) the feeder called “Sud” takes 1,469,172.6 kwh (see table 4.1) which represents 66.3% of the total annual energy loss. Because the peak current (of 200 ampere) demanded was very high therefore annual energy losses were also increased.

### D. Conclusion

The losses in 15-6.6 KV feeders (particularly in “Sud feeder”) is very high. The implication of those losses is that they reduce the Company profit, life of

cables and transformers, and the number of customers to be supplied with electricity. This has been hidden due the operational manual system in place. However, with a Computerized Web-Based Information System for computing losses, managers will have insight as to actual losses and be guided towards appropriate corrective and preventive maintenance necessary to minimize Goma distribution network losses.

#### E. Recommendation

From this study, the following recommendations becomes necessary for Goma electrical network distribution:

- Implementation of the Computerized Web-Based Information System in order to monitor losses over time for corrective and preventive maintenance of the electrical distribution network
- Immediate resizing of the overloaded “Sud feeder” or installation of another overhead cable that will take the half of the load in charge. Whenever the distribution transformers are overloaded and additional loads are anticipated, then the existing transformers should be replaced by higher capacity transformers, or new transformers may be provided to cater for the loads.

#### REFERENCES

- [1] B. G. Joseph, “ Use of computerized web-based information system for determining losses in traditional electrical network management: case study of Goma electrical distribution network”. MSc thesis, School of computer Studies, Kampala International University, xv+ 120 pp, April 2010

- [2] Global research Fund. “The Electricity

Economy: New opportunities from the transformation of the Electric Power Sector”, Global Smart energy, 2008.

- [3] W, Luke and T. Laura, PhP and MySQL Web Development, 2<sup>nd</sup>, Samms Publishing, Indianapolis .
- [4] A.S Pabla, Electric power distribution, 4<sup>th</sup> ed, McGraw-Hill , USA 2005.
- [5] Y, Wilis and D, David, “A General guide to writing research proposals and report, A hand book for beginning researchers, 2<sup>nd</sup> ed, Makerere University, Kampala, 2008.

#### AUTHORS PROFILE

**Ezekiel U. Okike** received the BSc degree in computer science from the University of Ibadan Nigeria in 1992, the Master of Information Science (MInfSc) in 1995 and PhD in computer science in 2007 all from the same University. He has been a lecturer in the Department of Computer Science, University of Ibadan since 1999 to date. Since September, 2008 to date, he has been on leave as a senior lecturer and Dean of the School of Computer Studies, Kampala International University, Uganda. His current research interests are in the areas of software engineering, software metrics, compilers and programming languages. He is a member of IEEE Computer and Communication societies.

**Bakunzi G. Joseph** is an Electrical Engineer from Institut Supérieur des Techniques Appliquées de Kinshasa and postgraduate student in the School of Computer Studies, Kampala International University . He holds the B.Eng. in Electrical Engineering and recently graduated with the degree of Master of Science in Systems Software Engineering , MSc(SSE).

# A DESIGN AND EXECUTION OF ACTIVITY-BASED APPLICATIONS IN DISTRIBUTED ENVIRONMENTS USING HERMES SOFTWARE ARCHITECTURE

**B. MUTHUKUMAR**

Assistant Professor  
Kathir College of Engineering,  
Coimbatore, Tamilnadu, INDIA  
muthu\_tcn@rediffmail.com

**P. BANUMATHI**

Assistant Professor  
Kathir College of Engineering,  
Coimbatore, Tamilnadu, INDIA  
banumathi\_mohankumar@yahoo.com

**T.K.P. RAJAGOPAL**

Assistant Professor  
Kathir College of Engineering,  
Coimbatore, Tamilnadu, INDIA  
tkprgrg@gmail.com

**R.VIDHYA**

PG Student, Pondicherry University, Pondicherry  
vidyaban@gmail.com

**Abstract** - Hermes is an agent-based middleware structured as component-based and 3-layered software architecture. Hermes provides an integrated, flexible programming environment for design and execution of activity-based applications in distributed environments. By using workflow technology, it supports even a non expert user programmer in the model driven design and implementation of a domain specific application. In this paper, after a description of Hermes software architecture, we provide a simple demo in biological domain and we show some real case studies in which Hermes has been validated.

**Keywords:** Hermes Software Architecture, O2I Project, Agents, Run –Time Layers etc.

## I. INTRODUCTION

Hermes [9] is an agent-based middleware, for design and execution of activity-based applications in distributed environments. It supports mobile computation as an application implementation strategy. While middleware for mobile computing has typically been developed to support physical and logical mobility, Hermes provides an integrated environment where application domain experts can focus on designing activity workflow and ignore the topological structure of the distributed environment. Generating mobile agents from a workflow specification is the responsibility of a context-aware compiler. Agents can also developed directly by an expert user using directly the Application

Programming Interface (API) provided by Hermes middleware. The Hermes middleware layer, compilers, libraries, services and other developed tools together result in a very general programming environment, which has been validated in two quite disparate application domains, one in industrial control [6] and the other in bioinformatics [13]. In the industrial control domain, embedded systems with scarce computational resources control product lines. Mobile agents are used to trace products and support self-healing. In the bioinformatics domain, mobile agents are used to support data collection and service discovery, and to simulate biological system through autonomous components interactions. This paper is organized as follows. Section II describes the Hermes Software Architecture. Section III provides a simple demo in biological domain. In Section IV, we present several projects in which Hermes middleware has been adopted. We conclude in Section V.

## II. HERMES SOFTWARE ARCHITECTURE

Hermes is structured as a component-based, agent oriented system with a 3-layer software architecture shown in Figure 1: user layer, system layer and run-time layer. At the user layer, it allows designers to specify their application as a workflow of activities using the graphical notation. At the system layer, it provides a context-aware compiler to generate a pool of user mobile agents from the workflow specification. At the run-time layer, it supports the activation of a set of specialized service agents, and it

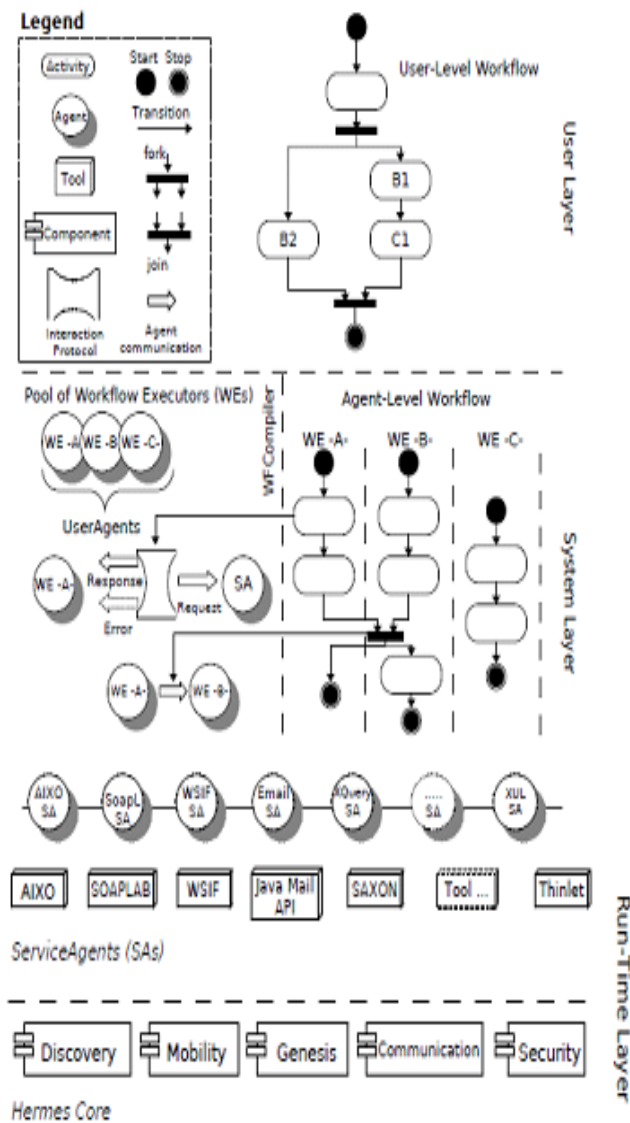


Fig. 1 Hermes Software Architecture

provides all necessary components to support agent mobility and communication.

The main difference between the run-time layer and the system layer is how agents function in each. Service Agents in the run-time layer are localized to one platform to interface with the local execution environment. User Agents in the system layer are workflow executors, created for a specific goal that, in theory, can be reached in a finite time by interacting with other agents. Afterwards that agent dies. Furthermore, for security UserAgents can access a local resource only by interacting with ServiceAgent that is the .guard. of the resource. It follows a detailed description of the main components and functionalities of each layer.

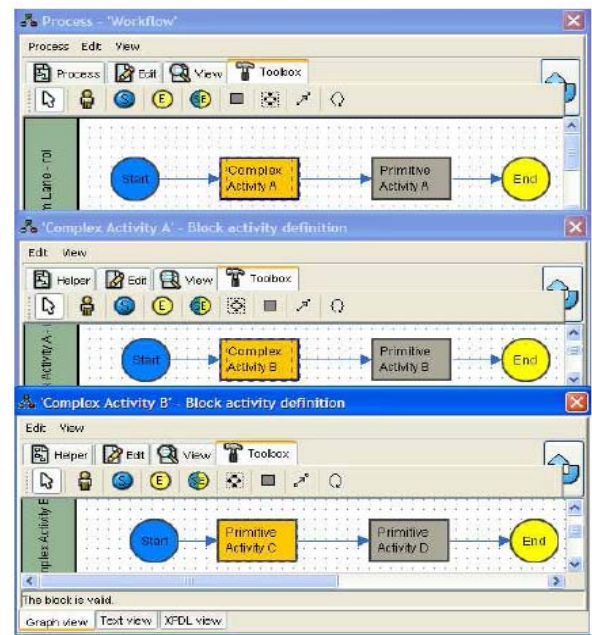


Fig. 2. Specification of Complex/Primitive activities in JaWE

## A. User Layer

The user layer is based on workflow technology and provides to users a set of programs for interacting with the workflow management system. There are two main families of programs: programs for specifying, managing and reusing existing workflow specifications, and programs enabling administration and direct interaction with the workflow management system. The workflow editor is the program that supports the workflows specification by composing activities in a graphical environment. Hermes provides two editors, one is a plugin of the stand-alone JaWE [10] editor and the other is WebWFlow, a web based editor. Both editors enable the specification of workflows by using XML Process Definition Language (XPDL) [14] a standard provided by the WfMC [12]. Activities used in a workflow are configured by specifying input parameters and their effects are recognizable as modification of state variables or modification on the environment's status. Workflow editors enable the composition of both primitive and complex activities.

A primitive activity is an activity that can be directly executed. Users can specify primitive activity without knowing the real implementation. A complex activity is an activity that must be specified before it can be used; as Figure 2 shows the specification of a complex activity could be a workflow of complex and/or simple activities. By using complex activities

the specification of workflows is simplified because they enhance both hierarchical specification and reuse: we can use an already existing complex activity without caring of its specification. Users can use complex activities and stored workflows to increase productivity when specifying new workflows. Moreover, large libraries of both domain specific primitives and complex activities can be loaded to specialize the editor for a specific application domain.

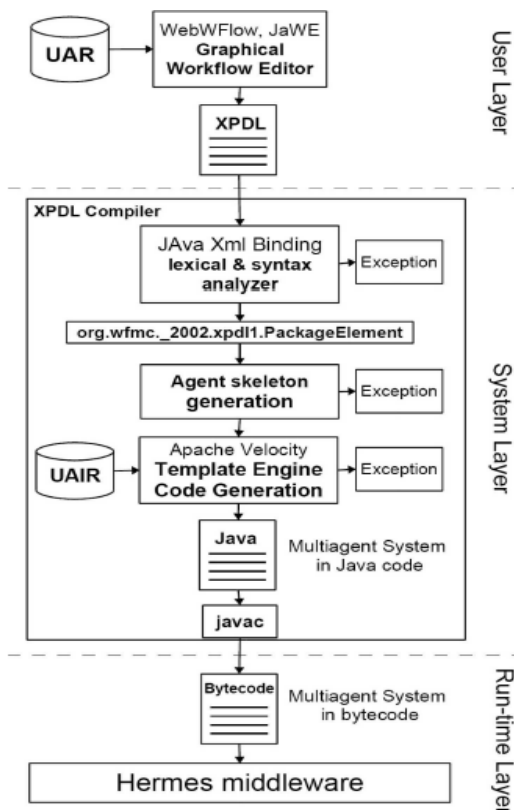


Fig. 3. Outline of workflow compilation process in Hermes

## B. System Layer

System Layer, on the middle architecture, provides the needed environment to map a user-level workflow into a set of primitive activities. The execution of these latter is coordinated by suitable model; they implement the activities at user level and embed implementation details abstracted from the execution environment. These primitive activities are implemented by autonomous software entities UserAgent able to react to the environment changes where they are executed. A compiler generates a pool of user mobile agents from the workflow specification. Due to the lack of space, workflow compilation process shown in Figure 3 will not discussed here and we refer to [4] for further details.

## C. Run-time Layer

Run-time Layer, at the bottom of the architecture, provides primitives and services essential for agent mobility and resources access. The kernel is the platform for mobile computing which provides primitives for discovery, mobility, communication, and security. As already described, the overall structure of the system is very complex, it supports abstract specifications that are mapped into a complex distributed and coordinated

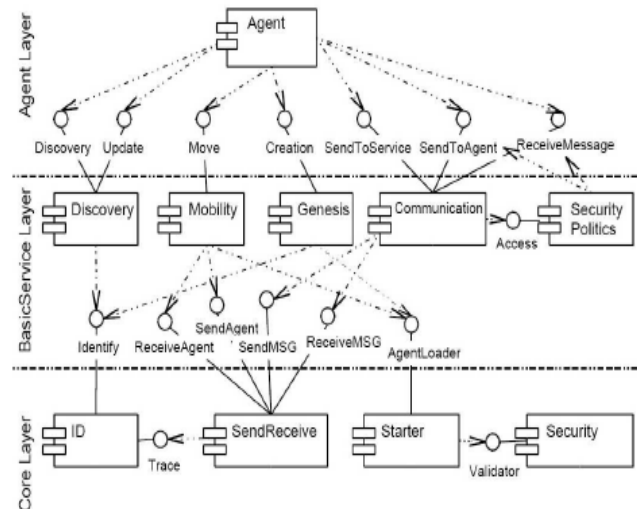


Fig. 4. 3-Layered Architecture of Hermes Mobile Computing Platform.

flows of activities over a large-scale distributed system. In order to master this complexity and to support the reusability of existing artefact during the development of a middleware system for a specific application domain, we designed Hermes kernel following a component-based [7] approach. Figure 4 shows the main components placed in the 3-Layered Architecture of Hermes Mobile Computing Platform. It follows a detailed description of components belonged to each layer.

**1) Core Layer:** It is the lowest layer of the architecture and contains base functions of the system, such as the implementation of the inter-platform communication protocols and agent management functions. This layer is composed of four components: ID, SendReceive, Starter and Security. The ID component, implements general identity management functions by managing a repository containing information about locally generated agents. This repository is accessed whenever we want to know the current position of an agent.



The ID component is also responsible for the creation of the identifiers to be associated to new agents. These identifiers contain information about birthplace, date and time of the agent's creation. Agent localization is simplified by information contained directly in the ID, such as birth place. In fact, the birth place of an agent hosts information about agent's current location. A second important feature of the Core is the SendReceive component.

This component implements low level inter-platform communication by sending and receiving messages and agents. By using traceability services offered by the ID component, SendReceive can easily update or retrieve the exact position of a specific user agent. The Starter component processes any request for agent creation. This particular component, in fact, take an inactive agent (just created or migrated), and checks it for the absence of malicious or manipulated code. These agents, before activation, are dynamically linked to all basic services of the platform. During execution the agent is isolated from the Core Layer by the BasicService layer. The Security component, as mentioned above, checks for the presence of malicious code or manipulations within agent code.

**2) BasicService Layer:** This layer has five main components: Discovery, Mobility, Genesis, Communication and Security Politics. The Discovery component searches and detects service agents. When a user agents wants to communicate with a service, it will ask the Discovery for the right identifier to use as the message's receiver. The service detection strategy can be implemented in several ways; for example by a fixed taxonomy or by an UDDI [5], commonly used in WebServices application domain. The mobility component enables the movement of code across platforms [11], it implements the interface used by the Agent component and it accesses to components of the Core layer to send, receive and load agents.

It is important to note that real communication between different locations can be achieved only through Core's SendReceive component, and then migration is independent of the type of used transport. Mobility consists on copy the agent i.e. its code and its current state and sends it to the destination platform where it will re-started in a specific point (weak mobility). The local agent is destroyed.

The Communication component makes possible to send and receive agent-directed messages both in an intra- and inter-platform context. Intra-platform messages are messages sent between agents

and services residing in the same platform. Inter-platform messages are messages sent to agents residing in different platforms (our system does not allow for remote communication between user agents and service agents).

The agent requesting the dispatch of a message does not need to know, effectively, where the target agent is; in fact, the ID is sufficient to post correctly a message. The Communication component uses one of the Security Policy's interfaces to ascertain whether the specific UserAgent or ServiceAgent has the right privileges for communication. If an Agent is not authorized to use a service, the message is destroyed. Before accessing resources and services, an agent must authenticate itself. The identification is performed by sending a login message to a specific ServiceAgent, as consequence the SecurityPolitics component jointly with the Communication component intercept the message and unlock the communication.

The SecurityPolitics component centralizes control of permissions, protects services and resources from the user agents, and provides the administrator with an easy way to manage all permissions. The last component of the service layer is the Genesis component that enables agent creation. A special case of agent creation is cloning that is performed when it is necessary to create a copy of an existing agent. The two copies differ only for the agent identifier.

**3) Agent Layer:** The Agent Layer is the upper layer of the mobile platform, the Agent Layer, contains all service and user agents. This component has not any interface, but it has only several dependencies upon the BasicService Layer. The Agent component provides a general abstract Agent class. UserAgent and UserAgent classes extend this abstract class. ServiceAgent consists of agents enabling access to local resources such data and tools.

User agents execute complex tasks and implement part of the logic of the application. Java programmers can also develop UserAgents by using the API provided by Hermes Mobile Computing Library. Listing 1 shows a simple demo. An MkDuckAgent called .Della Duck. Creates three sons Qui, Quo and Qua -lines 24 to 40- by cloning itself. After clonation each new agent starts its behavior calling "after Cloning" as initial method.

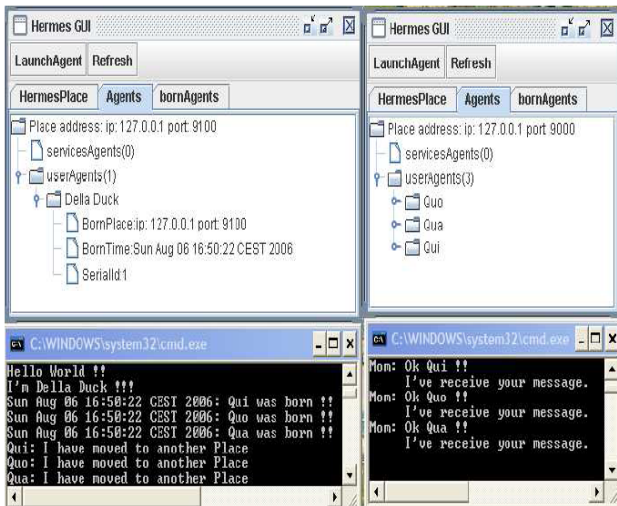


Fig. 5. Final result produced by McDuckAgent.java

```

1 package samples;
2 import hermesV2.*;
3 import hermesV2.agent.*;
4
5 public class McDuckAgent extends UserAgent {
6
7     public McDuckAgent(String agentName) {
8         super("Della Duck");
9     }
10
11     public void init() {
12         reception(); //I enable the reception
13                     //of messages for the father
14
15         System.out.println("Hello World !!");
16         System.out.println("I'm Della Duck !!!");
17
18         Identificator temp=null, son1=null,
19                     son2=null, son3=null;
20
21         /*afterCloning is the first method
22         called after the cloning*/
23
24         try {
25             son1 = clone("afterCloning", "Qui");
26             System.out.println(new Date(
27                 System.currentTimeMillis()) +
28                 ": Qui was born !!");
29             son2 = clone("afterCloning", "Quo");
30             System.out.println(new Date(
31                 System.currentTimeMillis()) +
32                 ": Quo was born !!");
33             son3 = clone("afterCloning", "Qua");
34
35             System.out.println(new Date(
36                 System.currentTimeMillis()) +
37                 ": Qua was born !!");
38         } catch (CloneException ce) {
39             System.out.println(ce);
40         }
41         Message m0=null, m1=null, m2=null, m3=null;
42         while (!(m1!=null && m2!=null &&
43             m3!=null)){
44             m0 = getMessageSynch();
45             temp = m0.getSenderAgentId();
46             if (son1.equals(temp)) m1 = m0;
47             if (son2.equals(temp)) m2 = m0;
48             if (son3.equals(temp)) m3 = m0;
49             System.out.println((String)m0.getObject());
50         }
51     }
52 }

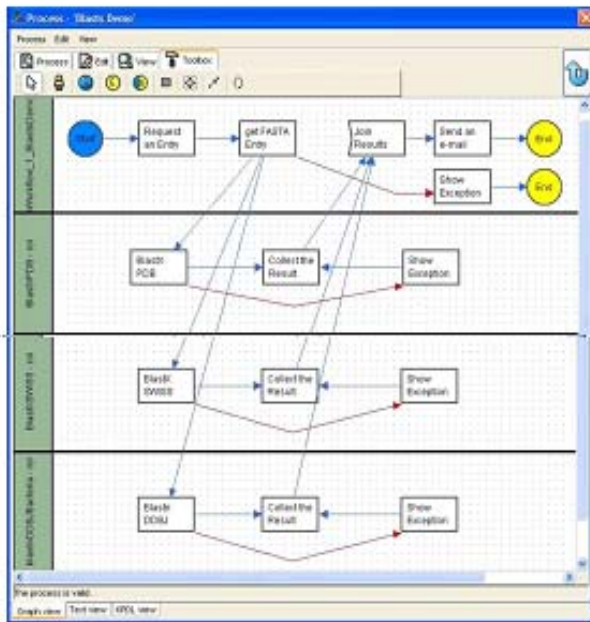
```

```

50 /*The mother replies to sons*/
51 Identificator myId = getIdentificator();
52 m1 = new Message(myId, son1,
53     "Mom: Ok Qui !! \n" +
54     "I've receive your message.");
55 m2 = new Message(myId, son2,
56     "Mom: Ok Quo !! \n" +
57     "I've receive your message.");
58 m3 = new Message(myId, son3,
59     "Mom: Ok Qua !! \n" +
60     "I've receive your message.");
61 try {
62     sendMessageToUserAgent(m1);
63     sendMessageToUserAgent(m2);
64     sendMessageToUserAgent(m3);
65 } catch (CommunicationException ce) {
66     System.out.println(ce.getMessage());
67 }
68 }
69
70 public void afterCloning() {
71     Identificator myId = getIdentificator();
72     PlaceAddress myBPA = myId.getBornPlaceAddress();
73     int myBPAPort = myBPA.getPort();
74     try {
75         int port = (myBPAPort == 9100) ? 9000 : 9100;
76
77         PlaceAddress myMPA =
78             new PlaceAddress(myBPA.getIp(), port);
79         this.move(myMPA, "afterMoving");
80     } catch (MigrationException me) {
81         System.out.println("MigrationException" + me);
82     }
83 }
84
85 public void afterMoving() {
86     reception(); //I enable the reception
87                 //of messages for the son
88
89     Identificator myId = getIdentificator();
90     Identificator mother = getFatherIdentificator();
91     Message m = null;
92
93     try {
94         m = new Message(myId, mother,
95             getAgentName() +
96             ": I have moved to another Place");
97         sendMessageToUserAgent(m);
98     } catch (CommunicationException ce) {
99         System.out.println(ce.getMessage());
100     }
101     m = getMessageSynch(mother);
102     System.out.println((String)m.getObject());
103 }
104 }

```

Listing 1. McDuckAgent.java



By using .move. method -line 79- Qui, Quo and Qua migrate to a Place different from where they were born. When they arrive in the new Place each one call the “afterMoving” -line 85- method. Then they notify to their mom their moving by using “sendMessageToUserAgent” -line 97- and “getMessageSynch” -line 101- methods. Figure 5 shows the final results.

#### D. Software requirements

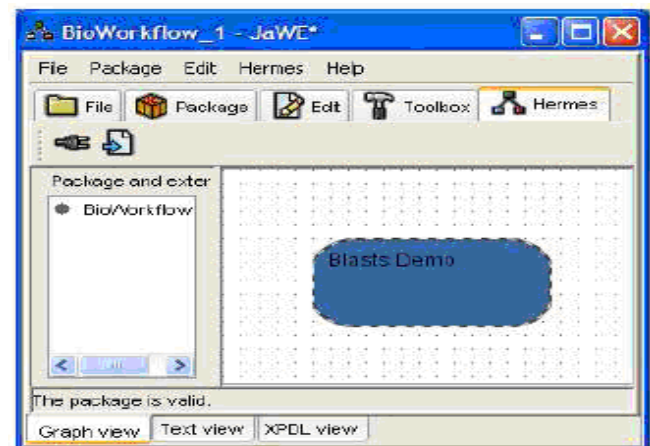
One of the main features of Hermes middleware is its scalability. The present version, HermesV2, is a pure Java application whose kernel requires about 120KB of memory and interoperates across a systems ranging from microprocessors to very power workstations. The Hermes Mobile Computing Platform is available under LGPL on Sourceforge 1 Web Site.

### III. MODEL DRIVEN DESIGN & IMPLEMENTATION OF ACTIVITY-BASED APPLICATIONS: A DEMO

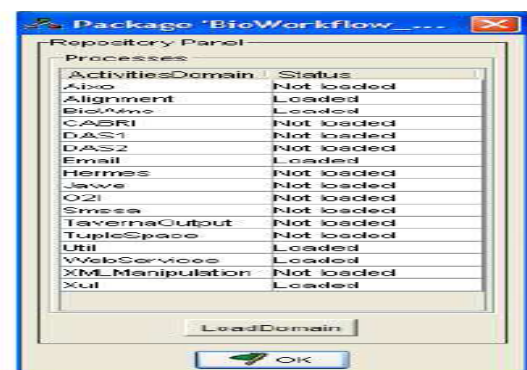
In the present post-genomic era, biological information sources are crammed with information gathered project from results of experiments performed in laboratories around the world, i.e., sequence alignments, hybridization data analysis or proteins interrelations. The amount of available information is constantly increasing, its wide distribution and the heterogeneity of the sources make difficult for bioscientists to manually collect and integrate information. In this section we present a

demo of Hermes in the biological domain. In our example we want to find similar DNA sequences to a given one in several databases using Basic Local Alignment Search Tool 2 (BLAST). In particular, in this demo we want to compare, using BLAST, the nucleotide sequence in FASTA format of a given entry identifier with the sequences contained in the following databases:

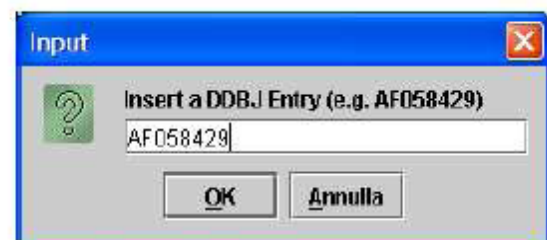
- Protein Data Bank (PDB) 3
- SWISS-PROT 4
- DDBJ 5



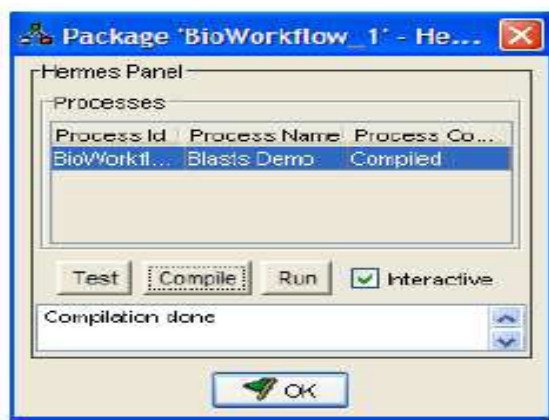
a)



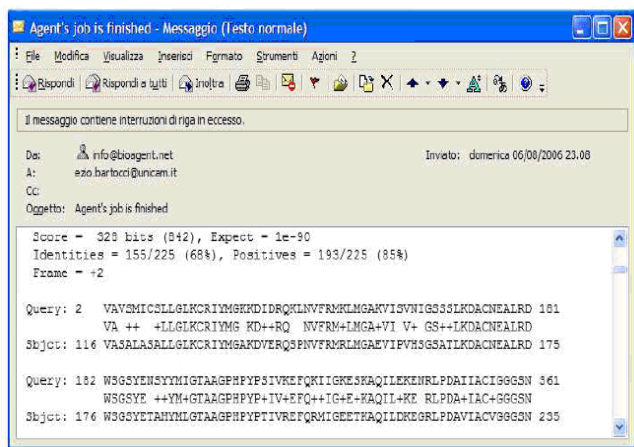
b)



c)



d)



e)

Fig. 7 Multiple blast compilation and execution

The access to these databases is guaranteed by a set of Web Services. By using workflow editors a bioscientists can specify the logic order, as Figure 6 shows, of a set of domain-specific activities without knowing the related implementation details. Each rectangle is an activity and each swimlane represents a UserAgent. As Figure 7-b shows, user can exploit a set of previous defined domain specific activities by importing the proper library. BlastnDDBJ Agent, BlastXSWISS Agent and BlastXPDB Agent receive the nucleotide sequence from the BlastDemo Agent and through an interaction with WSIF ServiceAgent; they compare the received sequence with sequences in each database using BLAST. If no exceptions occur, BlastDemo Agent join partial results and send the final document to user by email through an interaction with the Email ServiceAgent. After saving this specification, you can reload -Figure 7-a-, compile -Figure 7-d- and execute -Figure 7-c and 7-e – the workflow previous defined.

## IV. SOME CASE STUDIES

The Hermes middleware has been validated in several projects. It follows a brief case study description of Hermes application in some of them.

### A. SI.CO.M Project

In the SI.CO.M 6 project we have developed a prototype based on Hermes middleware for the traceability of ichthyic products. Generally the product is traced through the updating of databases distributed along the main sites of the weaving factory. This approach is not efficient because trace a faulty batch of products requires to query all databases, usually with an heterogeneous schema, of all sites interested in the production process. The proposed solution with the prototype named .TraceFish, exploiting the agent-based technology, allows moving automatically the information about a single batch from a site to another of the weaving factory overcoming the limits of the classical client/server approach.

### B. O2I Project

The Oncology over Internet (O2I) 7 project is aimed to develop a framework to support searching, retrieving and filtering information from Internet for oncology research and clinics. Hermes in the context of O2I project is called Bioagent 8, it supports the design and execution of user workflows involving access to literature, mutate on and cell lines databases.

### C. LITBIO Project

The main objective of the Laboratory of Interdisciplinary Technologies in Bioinformatics (LITBIO) 9 is to create infrastructure capable of supporting challenging international research and to develop new bioinformatics analysis strategies apply to biomedical and biotechnological data. To satisfy the most bioinformaticians needs we have proposed a multilayer architecture [2] based on Hermes middleware. At the user layer, it is intended to support in-silico experiments, resource discovery and biological systems simulation. The pivot of the architecture is a component called Resourceome [8], which keeps an alive index of resources in the bioinformatics domain using a specific ontology of resource information. A Workflow Management System, called BioWMS [3], provides a web-based interface to define in-silico experiments as workflows of complex and primitives activities. High level concepts concerning activities and data could be



indexed in the Resourceome that also dynamically supports workflow enactment, providing the related resources available at runtime. ORION [1], a multiagent system, is a proposed framework for modelling and engineering complex systems. The agent-oriented approach allows describing the behavior of the individual components and the rules governing their interactions. The agents also provide, as middleware, the necessary flexibility to support data and distributed applications. A GRID infrastructure allows a transparent access to the high performance computing resources required, for example in the biological systems simulation.

## V. CONCLUSION

As the demo presented shows, Hermes middleware provides an integrated, flexible programming environment, whose users can easily configure for its application domain. Hermes is structured as a component-based, agent oriented, 3-layered software architecture. It can configure for specific application domains by adding domain specific component libraries. The user can specify, modify and execute his workflow in a very simple way. Workflow is specified abstractly in a graphical notation and mapped to a set of autonomous computational units (UserAgents) interacting through a communication medium. The mapping is achieved by compiler that is aware not only of contents of a library of implemented user activities but also the software and hardware environment to executing them. By using workflow as suitable technology to hide distribution and on mobile agents as flexible implementation strategy of workflow in a distributed environment, Hermes allows even to a not expert programmer a model driven design and implementation of a domain specific activity-based application.

## REFERENCES

[1] M. Angeletti, A. Baldoncini, N. Cannata, F. Corradini, R. Culmone, C. Forcato, M. Mattioni, E. Merelli, and R. Piergallini. Orion: A spatial multi agent system framework for computational cellular dynamics of metabolic pathways. In Proceedings of Bioinformatics Italian Society (BITS) Meeting, Bologna, Italy, 2006.

[2] E. Bartocci, D. Cacciagrano, N. Cannata, F. Corradini, E. Merelli, and L. Milanesi. A GRID-based multilayer architecture for bioinformatics. In Proceedings of NETTAB'06 Network Tools and

Applications in Biology, Santa Margherita di Pula, Cagliari, Italy, 2006.

[3] E. Bartocci, F. Corradini, and E. Merelli. BioWMS: A web based workflow management system for bioinformatics. In Proceedings of Bioinformatics Italian Society (BITS) Meeting, Bologna, Italy, 2006.

[4] E. Bartocci, F. Corradini, and E. Merelli. Building a multiagent system from a user workflow specification. In Proceedings of Workshop From Objects to Agents - WOA, 2006.

[5] T. Bellwood, L. Clement, D. Ehnebuske, A. Hatley, M. Hondo, Y. L. Husband, K. Januszewski, S. Lee, B. McKee, J. Munter, and C. von Riegen. UDDI version 3.0. Published specification, Oasis, 2002.

[6] D. Bonura, F. Corradini, E. Merelli, and G. Romiti. Farmas: a MAS for extended quality workflow. In 2nd IEEE International Workshop on Theory and Practice of Open Computational Systems. IEEE Computer Society Press, 2004.

[7] D. Bonura, L. Mariani, and E. Merelli. Designing modular agent systems. In Proceedings of NET. Object DAYS, Erfurt, pages 245. 263, September 2003.

[8] N. Cannata, E. Merelli, and R. B. Altman. Time to organize the bioinformatics resourceome. PLoS Comput Biol., 1(7):e76, 2005.

[9] F. Corradini and E. Merelli. Hermes: agent-based middleware for mobile computing. In Mobile Computing, volume 3465, pages 234.270. LNCS, 2005.

[10] Enhydra. Jawe. <http://jawe.enhydra.org/>, 2003.

[11] A. Fuggetta, G. Picco, and G. Vigna. Understanding code mobility. IEEE Transaction of Software Engineering, 24(5):352.361, May 1998.

[12] D. Hollingsworth. The Workflow Reference Model, January 1995.

[13] E. Merelli, R. Culmone, and L. Mariani. Bioagent: a mobile agent system for bioscientists. In NETTAB Workshop on Agents and Bioinformatics, Bologna, July 2002.

[14] WfMC. Xml process definition language (xpd). WfMC standard, W3C, October 2005.

## AUTHORS PROFILE



**B.MUTHUKUMAR** received B.E (CSE) degree from Anna University, Chennai in the year 2005. M.Tech (CSE) with Gold Medal from Dr. MGR. University in the year 2007, and Distinction with Honors. Now Pursuing Ph.D in St. Peter's University, Chennai from 2008. Having 6 Years of Teaching Experience. His area of interest in Image Processing and Networks. He published 04 International Conference and 15 National Conference and 28 National Workshops / FDP / Seminars etc., He is a member of ISTE, CSI, IEEE, Member of IACSIT and Member of IAENG.



**P.BANUMATHI** received B.E, MCA, MBA and M.Phil in the year 1994, 2003, 2008 and 2007. She is having 8 Years of teaching experience and 5 years of Industrial experience. Her area of interest is Neural Networks. She published 01 International Conference and 10 National Conference. She is a member of IEEE, CSI, Member of IACSIT and Member of IAENG.



**T.K.P. RAJAGOPAL** M.C.A., M.Phil., M.A., M.Phil., M.E.(CSE), M.Tech., is working as an Assistant Professor in the Department of Computer Science and Engineering at Kathir College of Engineering, Coimbatore, Tamilnadu, India. He has 11 years of teaching experience. His research areas are Network Security, Data Mining and Digital Image Processing. He has guided more than 100 M.Phil, Computer Science

students of various Universities. He has published 3 books, 2 research papers in international journals, Presented 2 papers in International Conference, 13 papers in National level conferences and has attended 12 Workshops. He is a life member of various professional societies like ISTE, DOEACC, CSTA, ACS, IAENG and IACSIT.



# TRACKING THE FACE MOVEMENT OF PERCEPTUAL USER INTERFACE USING CAMSHIFT ALGORITHM

B. MUTHUKUMAR<sup>1</sup>

Assistant Professor, Information Technology,  
Kathir College of Engineering, Coimbatore – 62  
Email: [muthu\\_tcn@rediffmail.com](mailto:muthu_tcn@rediffmail.com)

Dr.S.RAVI<sup>2</sup>

Professor and Head, ECE Department,  
Dr.M.G.R University, Chennai – 95  
Email: [ravi\\_mls@yahoo.com](mailto:ravi_mls@yahoo.com)

**Abstract:** This paper deals with a perceptual user interface and computer vision color tracking algorithm is developed and applied towards tracking human faces. Computer vision algorithms that are intended to form part of a perceptual user interface must be fast and efficient. They must be able to track in real time yet not absorb a major share of computational resources: other tasks must be able to run while the visual interface is being used. The new algorithm developed here is based on a robust nonparametric technique for climbing density gradients to find the mode (peak) of probability distributions called the mean shift algorithm. In our case, we want to find the mode of a color distribution within a video scene. Therefore, the mean shift algorithm is modified to deal with dynamically changing color probability distributions derived from video frame sequences. The modified algorithm is called the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm. CAMSHIFT's tracking accuracy is compared against a Polhemus tracker. Tolerance to noise, distractors and performance is studied.

**Keywords:** Computer vision, Face tracking, Mean Shift Algorithm, Perceptual User Interface, 3D Graphics Interface

## 1. INTRODUCTION

Perceptual interfaces are ones in which the computer is given the ability to sense and produce analogs of the human senses, such as allowing computers to perceive and produce localized sound and speech, giving computers a sense of touch and force feedback, and in our case, giving computers an ability to see. Computer vision face tracking is an active and developing field, yet the face trackers that have been developed are not sufficient for our needs. Elaborate methods such as tracking contours with snakes [[10][12][13]], using Eigenspace matching techniques [14], maintaining large sets of statistical hypotheses [15], or convolving images with feature detectors [16] are far too computationally expensive. We want a tracker that will track a given face in the presence of noise, other faces, and hand movements.

Moreover, it must run fast and efficiently so that objects may be tracked in real time (30 frames per second) while consuming as few system resources as possible. The mean shift algorithm operates on probability distributions. To track colored objects in video frame sequences, the color image data has to be represented as a probability distribution [1]; we use color histograms to accomplish this. Color distributions derived from video image sequences change over time, so the mean shift algorithm has to be modified to adapt dynamically to the probability distribution it is tracking. The new algorithm that meets all these requirements is called CAMSHIFT.

For face tracking, CAMSHIFT tracks the X, Y, and Area of the flesh color probability distribution representing a face. Area is proportional to Z, the distance from the camera. Head roll is also tracked as a further degree of freedom. We then use the X, Y, Z, and Roll derived from CAMSHIFT face tracking as a perceptual user interface for controlling commercial computer games and for exploring 3D

graphic virtual worlds. Figure 1 summarizes the algorithm described below. For each video frame, the raw image is converted to a color probability distribution image via a color histogram model of the color being tracked (flesh for face tracking).

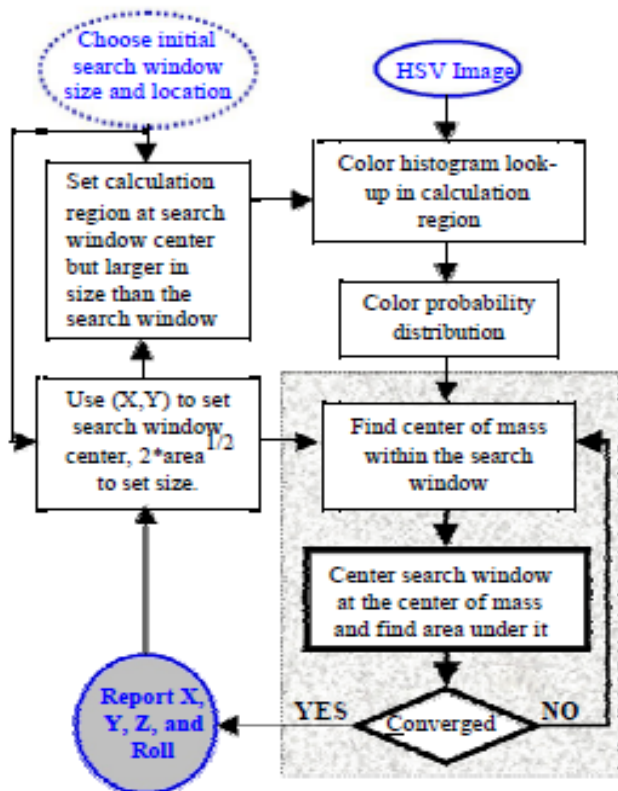


Figure 1: Block diagram of color object tracking

The center and size of the color object are found via the CAMSHIFT algorithm operating on the color probability image (the gray box is the mean shift algorithm). The current size and location of the tracked object are reported and used to set the size and location of the search window in the next video image. The process is then repeated for continuous tracking.

## 2. VIDEO DEMONSTRATIONS

The following three videos demonstrate CAMSHIFT in action.

1. FaceTrack\_Fast.avi
2. FaceTrack\_Distractors.avi
3. FaceTrack\_HandOcclusion.avi

The first video shows CAMSHIFT tracking rapid face movements. The second video shows CAMSHIFT tracking a face with other faces moving in the scene. The third video shows CAMSHIFT tracking a face through hand occlusions.

### 2.1 Color Probability Distributions

In order to use CAMSHIFT to track colored objects in a video scene, a probability distribution image of the desired color (flesh color in the case of face tracking) in the video scene must be created. In

order to do this, we first create a model of the desired hue using a color histogram. We use the Hue Saturation Value (HSV) color system [5][6] that corresponds to projecting standard Red, Green, Blue (RGB) color space along its principle diagonal from white to black (see arrow in Figure 2). This results in the hexcone in Figure 3. Descending the V axis in Figure 3 gives us smaller hexcones corresponding to smaller (darker) RGB sub cubes in Figure 2. HSV space separates out hue (color) from saturation (how concentrated the color is) and from brightness. We create our color models by taking 1D histograms from the H (hue) channel in HSV space. For face tracking via a flesh color model, flesh areas from the user are sampled by prompting users to center their face in an onscreen box, or by using motion cues to find flesh areas from which to sample colors.

The hues derived from flesh pixels in the image are sampled from the H channel and binned into a 1D histogram. When sampling is complete, the histogram is saved for future use. More robust histograms may be made by sampling flesh hues from multiple people. Even simple flesh histograms tend to work well with a wide variety of people without having to be updated. A common misconception is that different color models are needed for different races of people, for example, for blacks and whites. This is not true. Except for albinos, humans are all the same color (hue). Darkskinned people simply have greater flesh color saturation than light-skinned people, and this is separated out in the HSV color system and ignored in our flesh-tracking color model.

During operation, the stored flesh color histogram is used as a model, or lookup table, to convert incoming video pixels to a corresponding probability of flesh image as can be seen in the right-hand image of Figure 6. This is done for each video frame. Using this method, probabilities range in discrete steps from zero (probability 0.0) to the maximum probability pixel value (probability 1.0). For 8-bit hues, this range is between 0 and 255. We then track using CAMSHIFT on this probability of flesh image. When using real cameras with discrete pixel values, a problem can occur when using HSV space as can be seen in Figure 3.

When brightness is low (V near 0), saturation is also low (S near 0). Hue then becomes quite noisy, since in such a small hexcone, the small number of discrete hue pixels cannot adequately represent slight changes in RGB. This then leads to wild swings in hue values. To overcome this problem, we simply ignore hue pixels that have very low corresponding brightness values. This means that for very dim scenes, the

camera must auto-adjust or be adjusted for more brightness or else it simply cannot track. With sunlight, bright white colors can take on a flesh hue so we also use an upper threshold to ignore flesh hue pixels with corresponding high brightness. At very low saturation, hue is not defined so we also ignore hue pixels that have very low corresponding saturation. Originally, we used a 2D color histogram built from normalized red green (r, g) space ( $r = R/(R+G+B)$ ,  $g = G/(R+G+B)$ ). However, we found that such color models are much more sensitive to lighting changes since saturation (which is influenced by lighting) is not separated out of that model.

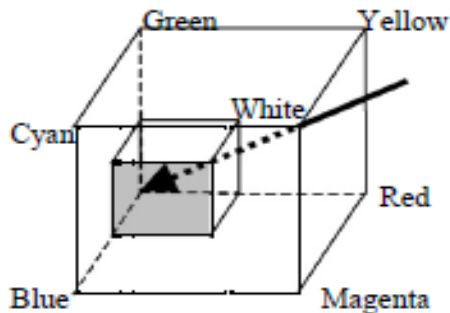


Figure 2: RGB color cube

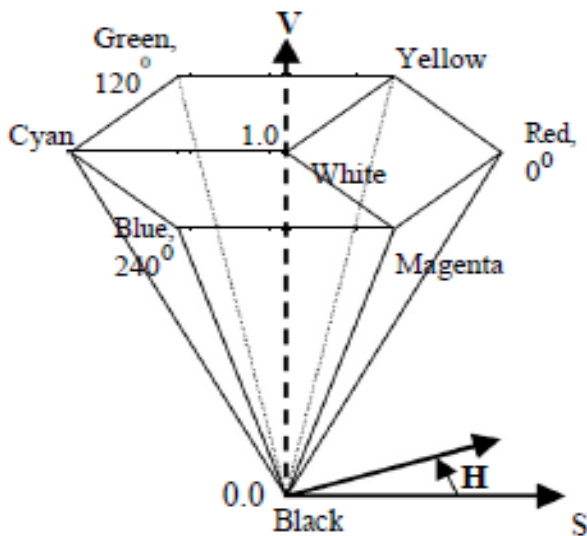


Figure 3: HSV color system

### 3. CAMSHIFT DERIVATION

The closest existing algorithm to CAMSHIFT is known as the mean shift algorithm [2][18]. The mean shift algorithm is a non-parametric technique that climbs the gradient of a probability distribution to find the nearest dominant mode (peak).

#### 3.1 How to Calculate the Mean Shift Algorithm

1. Choose a search window size.
2. Choose the initial location of the search window.

3. Compute the mean location in the search window.
4. Center the search window at the mean location computed in Step 3.
5. Repeat Steps 3 and 4 until convergence (or until the mean location moves less than a preset threshold).

Unlike the Mean Shift algorithm, which is designed for static distributions, CAMSHIFT is designed for dynamically changing distributions. These occur when objects in video sequences are being tracked and the object moves so that the size and location of the probability distribution changes in time. The CAMSHIFT algorithm adjusts the search window size in the course of its operation. Initial window size can be set at any reasonable value. For discrete distributions (digital data), the minimum window size is three as explained in the Implementation Details section.

Instead of a set or externally adapted window size, CAMSHIFT relies on the zeroth moment information, extracted as part of the internal workings of the algorithm, to continuously adapt its window size within or over each video frame. One can think of the zeroth moment as the distribution “area” found under the search window. Thus, window radius, or height and width, is set to a function of the zeroth moment found during search. The CAMSHIFT algorithm is then calculated using any initial non-zero window size (greater or equal to three if the distribution is discrete).

#### 3.2 How to Calculate the Continuously Adaptive Mean Shift Algorithm

1. Choose the initial location of the search window.
2. Mean Shift as above (one or many iterations); store the zeroth moment.
3. Set the search window size equal to a function of the zeroth moment found in Step 2.
4. Repeat Steps 2 and 3 until convergence (mean location moves less than a preset threshold).

In Figure 4 below, CAMSHIFT is shown beginning the search process at the top left step by step down the left then right columns until convergence at bottom right. In this figure, the red graph is a 1D cross-section of an actual sub-sampled flesh color probability distribution of an image of a face and a nearby hand. In this figure, yellow is the CAMSHIFT search window, and purple is the mean shift point. The ordinate is the distribution value, and the abscissa is the horizontal spatial position within the original image. The window is initialized at size three and converges to cover the tracked face but not the hand in six



iterations. In this sub-sampled image, the maximum distribution pixel value is 206 so we set the width of the search window to be  $2 * M_0 / 206$  (see discussion of window size in the Implementation Details section below). In this process, CAMSHIFT exhibits typical behavior: it finds the center of the nearest connected distribution region (the face), but ignores nearby distractors (the hand).

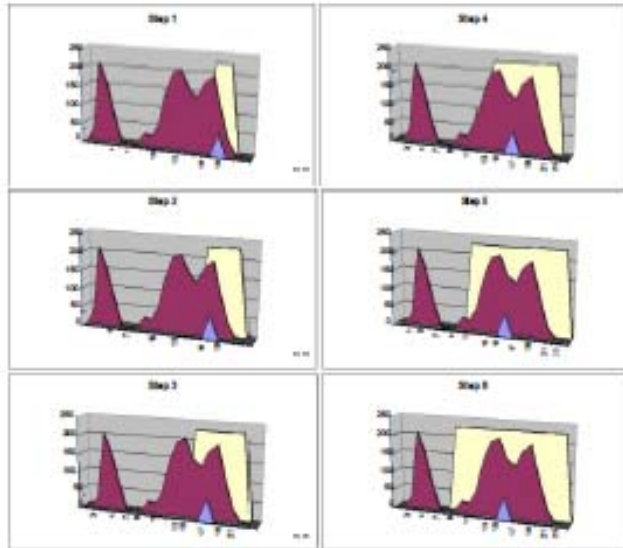


Figure 4: CAMSHIFT in operation down the left then right columns

Figure 4 shows CAMSHIFT at startup. Figure 5 below shows frame to frame tracking. In this figure, the red color probability distribution has shifted left and changed form. At the left in Figure 5, the search window starts at its previous location from the bottom right in Figure 4. In one iteration it converges to the new face center.

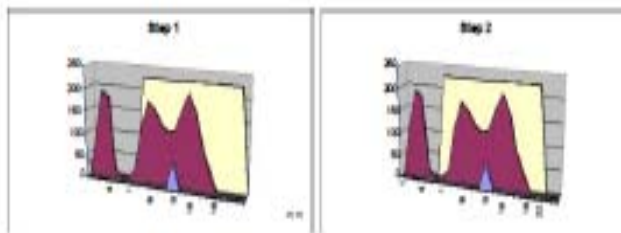


Figure 5: Example of CAMSHIFT tracking starting from the converged search location in Figure 4 bottom right

### 3.3 Mean Shift Alone Does Not Work

The mean shift algorithm alone would fail as a tracker. A window size that works at one distribution scale is not suitable for another scale as the color object moves towards and away from the camera. Small fixed-sized windows may get lost entirely for large object translation in the scene. Large fixed-sized windows may include distractors (other people or hands) and too much noise.

## 4. CAMSHIFT FOR VIDEO SEQUENCES

When tracking a colored object, CAMSHIFT operates on a color probability distribution image derived from color histograms. CAMSHIFT calculates the centroid of the 2D color probability distribution within its 2D window of calculation, re-centers the window, and then calculates the area for the next window size.

Thus, we needn't calculate the color probability distribution over the whole image, but can instead restrict the calculation of the distribution to a smaller image region surrounding the current CAMSHIFT window. This tends to result in large computational savings when flesh color does not dominate the image. We refer to this feedback of calculation region size as the Coupled CAMSHIFT algorithm.

### 4.1 How to Calculate the Coupled CAMSHIFT Algorithm

1. First, set the calculation region of the probability distribution to the whole image.
2. Choose the initial location of the 2D mean shift search window.
3. Calculate the color probability distribution in the 2D region centered at the search window location in an area slightly larger than the mean shift window size.
4. Mean shift to convergence or for a set number of iterations. Store the zeroth moment (area or size) and mean location.
5. For the next video frame, center the search window at the mean location stored in Step 4 and set the window size to a function of the zeroth moment found there. Go to Step 3.

For each frame, the mean shift algorithm will tend to converge to the mode of the distribution. Therefore, CAMSHIFT for video will tend to track the center (mode) of color objects moving in a video scene. Figure 6 shows CAMSHIFT locked onto the mode of a flesh color probability distribution (mode center and area are marked on the original video image). In this figure, CAMSHIFT marks the face centroid with a cross and displays its search window with a box.



Figure 6: A video image and its flesh probability image

#### 4.2 How CAMSHIFT Deals with Image Problems

When tracking color objects, CAMSHIFT deals with the image problems mentioned previously of irregular object motion due to perspective, image noise, distractors, and facial occlusion as described below. CAMSHIFT continuously re-scales itself in a way that naturally fits the structure of the data. A colored object's potential velocity and acceleration scale with its distance to the camera, which in turn, scales the size of its color distribution in the image plane. Thus, when objects are close, they can move rapidly in the image plane, but their probability distribution also occupies a large area.

In this situation, CAMSHIFT's window size is also large and so can catch large movements. When objects are distant, the color distribution is small so CAMSHIFT's window size is small, but distal objects are slower to traverse the video scene. This natural adaptation to distribution scale and translation allows us to do without predictive filters or variables—a further computational saving—and serves as an in-built antidote to the problem of erratic object motion. CAMSHIFT's windowed distribution gradient climbing causes it to ignore distribution outliers. Therefore, CAMSHIFT produces very little jitter in noise and, as a result, tracking variables do not have to be smoothed or filtered. This gives us robust noise tolerance.

CAMSHIFT's robust ability to ignore outliers also allows it to be robust against distractors. Once CAMSHIFT is locked onto the mode of a color distribution, it will tend to ignore other nearby but non-connected color distributions. Thus, when CAMSHIFT is tracking a face, the presence of other faces or hand movements in the scene will not cause CAMSHIFT to lose the original face unless the other faces or hand movements substantially occlude the original face. CAMSHIFT's provable convergence to the mode of probability distributions helps it ignore partial occlusions of the colored object. CAMSHIFT will tend to stick to the mode of the color distribution that remains.

Moreover, when CAMSHIFT's window size is set somewhat greater than the root of the distribution area under its window, CAMSHIFT tends to grow to encompass the connected area of the distribution that is being tracked (see Figure 4). This is just what is desired for tracking whole objects such as faces, hands, and colored tools. This property enables CAMSHIFT to not get stuck tracking, for example, the nose of a face, but instead to track the whole face.

### 5. IMPLEMENTATION DETAILS

#### 5.1 Initial Window Size and Placement

In practice, we work with digital video images so our distributions are discrete. Since CAMSHIFT is an algorithm that climbs the gradient of a distribution,

the minimum search window size must be greater than one in order to detect a gradient. Also, in order to center the window, it should be of odd size. Thus for discrete distributions, the minimum window size is set at three. For this reason too, as CAMSHIFT adapts its search window size, the size of the search window is rounded up to the current or next greatest odd number. In practice, at start up, we calculate the color probability of the whole scene and use the zeroth moment to set the window size (see subsection below) and the centroid to set the window center.

#### 5.2 Setting Adaptive Window Size Function

Deciding what function of the zeroth moment to set the search window size to in Step 3 of the CAMSHIFT algorithm depends on an understanding of the distribution that one wants to track and the goal that one wants to achieve. The first consideration is to translate the zeroth moment information into units that make sense for setting window size. Thus, in Figure 4, the maximum distribution value per discrete cell is 206, so we divide the zeroth moment by 206 to convert the calculated area under the search window to units of number of cells.

Our goal is then to track the whole color object so we need an expansive window. Thus, we further multiply the result by two so that the window grows to encompass the connected distribution area. We then round to the next greatest odd search window size so that the window has a center.

#### 5.3 Comments on Software Calibration

Much of CAMSHIFT's robustness to noise, transient occlusions, and distractors depends on the search window matching the size of the object being tracked—it is better to err on the side of the search window being a little too small.

The search window size depends on the function of the zeroth moment  $M_{00}$  chosen above. To indirectly control the search window size, we adjust the color histogram up or down by a constant, truncating at zero or saturating at the maximum pixel value. This adjustment affects the pixel values in the color probability distribution image which affects  $M_{00}$  and hence window size. For 8-bit hue, we adjust the histogram down by 20 to 80 (out of a maximum of 255), which tends to shrink the CAMSHIFT window to just within the object being tracked and also reduces image noise.

#### 5.4 Comments on Hardware Calibration

To use CAMSHIFT as a video color object tracker, the camera's field of view (zoom) must be set so that it covers the space that one intends to track in. Turn off automatic white balance if possible to avoid sudden color shifts. Try to set (or auto-adjust) AGC, shutter speed, iris or CCD

integration time so that image brightness is neither too dim nor saturating. The camera need not be in focus to track colors. CAMSHIFT will work well with cheap cameras and does not need calibrated lenses.

#### 5.5 CAMSHIFT's Actual Use as an Interface

CAMSHIFT is being used as a face tracker to control games and 3D graphics. By inserting face control variables into the mouse queue, we can control unmodified commercial games such as Quake 2 shown in Figure 7. We used left and right head movements to slide a user left and right in the game, back and forth head movements to move the user backwards and forwards, up or down movements to let the user shoot (as if ducking or getting jolted by the gun), and roll left or right to turn the user left or right in the game. This methodology has been used extensively in a series of demos with over 30 different users. Head tracking via CAMSHIFT has also been used to experiment with immersive 3D graphics control in which natural head movements are translated to moving the corresponding 3D graphics camera viewpoint.

This has been extensively tested using a 3D graphics model of the Forbidden City in China as well as in exploring a 3D graphics model of the big island of Hawaii as shown in Figure 8. Most users find it an enjoyable experience in which they naturally pick up how to control the graphics viewpoint movement.



Figure 7: CAMSHIFT-based face tracker used to play Quake 2 hands free by inserting control variables into the mouse queue

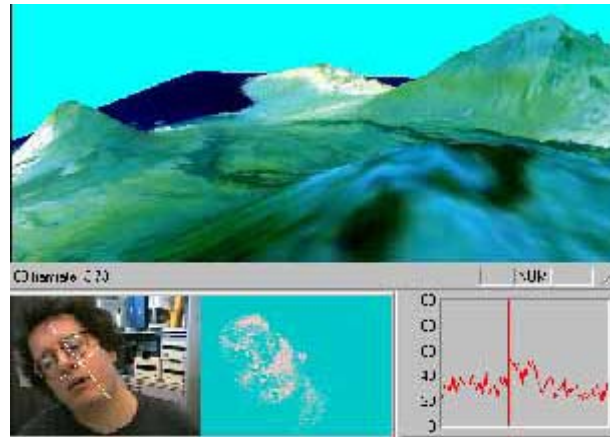


Figure 8: CAMSHIFT-based face tracker used to “fly” over 3D graphic’s model of Hawaii

## 6. CAMSHIFT ANALYSIS

### 6.1 Comparison to Polhemus

In order to assess the tracking accuracy of CAMSHIFT, we compared its accuracy against a Polhemus tracker. Polhemus is a magnetic sensor connected to a system that measures six degrees of spatial freedom and thus can be used for object tracking when tethered to an object.

The observed accuracy of Polhemus is  $\pm 1.5\text{cm}$  in spatial location and about  $2.5^\circ$  in orientation within 30 inches of the Polhemus antenna. We compared Polhemus tracking to CAMSHIFT color object tracking using a  $320 \times 240$  image size (see Figure 14a-d). The coordinate systems of Polhemus and the camera were carefully aligned prior to testing.

The object tracked was pulled on a cart in a set trajectory away from the Polhemus origin. The comparison between CAMSHIFT and Polhemus in each of X, Y, Z, and Roll yielded the results shown Table 1.

Tracking Variable	X	Y	Z	Roll
Standard Deviation of Difference	0.27cm	0.58cm	3.4cm	$2.4^\circ$

Table 1: Standard deviation of Polhemus vs. CAMSHIFT tracking differences



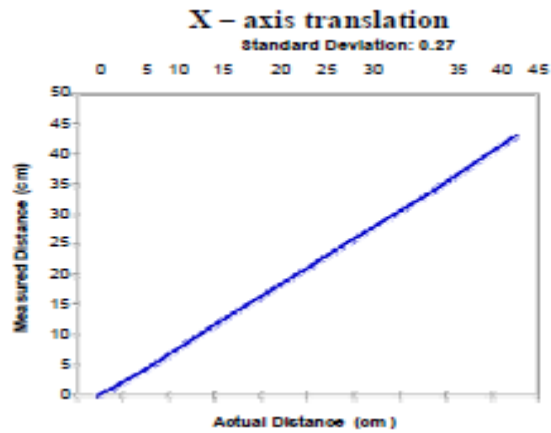


Figure 14a: Comparison of X tracking accuracy

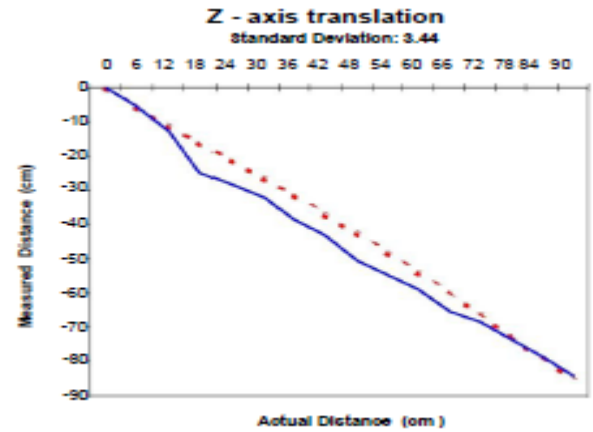


Figure 14c: Comparison of Z tracking accuracy

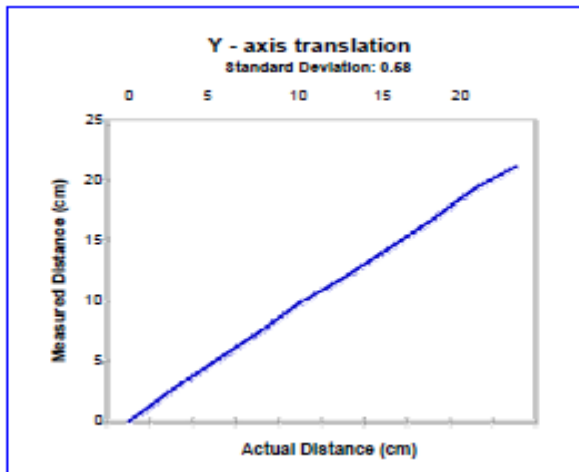


Figure 14b: Comparison of Y tracking accuracy

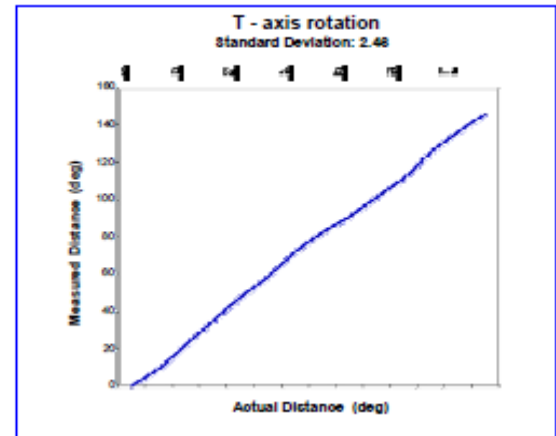


Figure 14d: Accuracy comparison of Polhemus and CAMSHIFT tracking for roll.



## 6.2 Tracking In Noise

CAMSHIFT's robust ability to find and track the mode of a dynamically changing probability distribution also gives it good tracking behavior in noise. We videotaped a head movement sequence and then played it back adding 0, 10, 30, and 50% uniform noise. Figure 15 shows 50% noise added to the raw image on the left, and the resulting color probability distribution on the right. Note that the use of a color model greatly cuts down the random noise since color noise has a low probability of being flesh color.

Nevertheless, the flesh color model is highly degraded and there are many spurious flesh pixels in the color probability distribution image. But CAMSHIFT is still able to track X, Y, and Roll quite well in up to 30% white noise as shown in Figure 16a-d. Z is more of a problem because CAMSHIFT measures Z by tracking distribution area under its search window, and one can see in Figure 9 that area is highly affected by noise. Y shows an upward shift simply because the narrower chin region exhibits more degradation in noise than the wider forehead. Roll tracks well until noise is such that the length and width of the face color distribution are obscured. Thus, CAMSHIFT handles noise well without the need for extra filtering or adaptive smoothing.



Figure 9: Tracking in 50% uniform noise

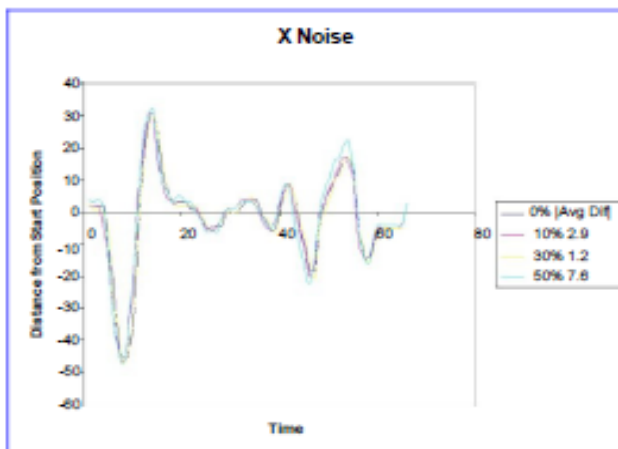


Figure 9a: X accuracy in 0-50% uniform noise

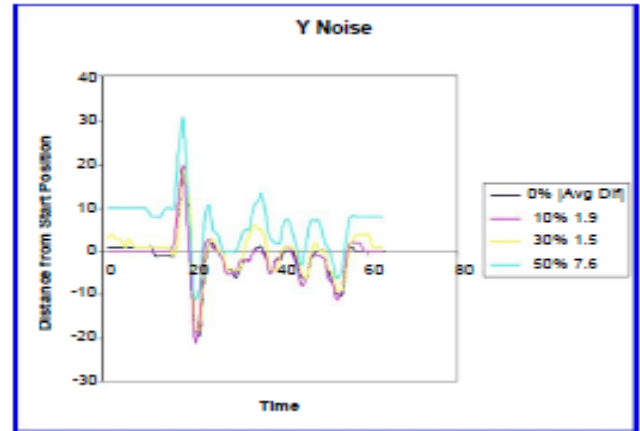


Figure 9b: Y accuracy in 0-50% uniform noise

## 7. DISCUSSION

This paper discussed a core tracking module that is part of a larger effort to allow computers to track and understand human motion, pose, and tool use. As such, the module was designed to be simple and computationally efficient. Yet, this core module must still handle the basic computer-vision problems outlined in this paper. We've seen that CAMSHIFT handles these problems as follows:

### 7.1 Irregular object motion:

CAMSHIFT scales its search window to object size thus naturally handling perspective-induced motion irregularities.

### 7.2 Image noise:

The color model eliminates much of the noise, and CAMSHIFT tends to ignore the remaining outliers.

### 7.3 Distractors:

CAMSHIFT ignores objects outside its search window so objects such as nearby faces and hands do not affect CAMSHIFT's tracking.

### 7.4 Occlusion:

As long as occlusion isn't 100%, CAMSHIFT will still tend to follow what is left of the objects' probability distribution.

### 7.5 Lighting variation:

Using only hue from the HSV color space and ignoring pixels with high or low brightness gives CAMSHIFT wide lighting tolerance. CAMSHIFT's simplicity does cause limitations however. Since CAMSHIFT derives Z from object area estimates, Z is subject to noise and spurious values. The effects of noise are evident in Figure 9c. Since CAMSHIFT relies on color distributions alone, errors in color (colored lighting, dim illumination, too much illumination) will cause errors in tracking. More sophisticated trackers use multiple modes such as

feature tracking and motion analysis to compensate for this, but more complexity would undermine the original design criterion for CAMSHIFT.

CAMSHIFT also only detects four (X, Y, Z, and Roll) of the six modes of freedom (above plus pitch and yaw). Unfortunately, of the six degrees of head movement possible, Roll is the least useful control variable since it is the least "natural" head movement and is therefore fatiguing for the user to use constantly.

## 8. CONCLUSION

CAMSHIFT is a simple, computationally efficient face and colored object tracker. While acknowledging the limitation imposed by its simplicity, we can still see that CAMSHIFT tracks virtually as well as more expensive tethered trackers (Polhemus) or much more sophisticated, computationally expensive vision systems and it tracks well in noisy environments. Thus, as we have shown, even though CAMSHIFT was conceived as a simple part of a larger tracking system, it has many uses right now in game and 3D graphics' control. Adding perceptual interfaces can make computers more natural to use, more fun for games and graphics, and a better medium of communication. These new features consume more MIPs and so will take advantage of more MIPs available with future Intel® CPUs. In this project, we designed a highly efficient face tracking algorithm rather than a more complex, higher MIPs usage algorithm. This was done because we want to be able to demonstrate compelling applications and interfaces on today's systems in order to prepare the way for the future use of computer vision on PCs. CAMSHIFT is usable as a visual interface now, yet designed to be part of a more robust, larger tracking system in the future. CAMSHIFT will be incorporated into larger, more complex, higher MIPs-demanding modules that provide more robust tracking, posture understanding, gesture and face recognition, and object understanding. In this way, the functionality of the computer vision interface will increase with increasing Intel CPU speeds.

## REFERENCES

- [1] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," CVPR'97, pp. 750-755.
- [2] K. Fukunaga, "Introduction to Statistical Pattern Recognition," Academic Press, Boston, 1990.
- [3] MMXTM technology optimized libraries in image, signal processing, pattern recognition and matrix math can be

mdownloadedfrom<http://developer.intel.com/design/perftool/perflibst/index.htm>

- [4] W.T. Freeman, K. Tanaka, J.Ohta, and K. Kyuma, "Computer Vision for Computer Games," Int. Conf. On Automatic Face and Gesture Recognition, pp. 100-105, 1996.
- [5] A.R. Smith, "Color Gamut Transform Pairs," SIGGRAPH 78, pp. 12-19, 1978.
- [6] J.D. Foley, A. van Dam, S. K. Feiner and J.F. Hughes, "Computer graphics principles and practice," Addison-Wesley, pp. 590-591.
- [7] P. Fieguth and D. Terzopoulos, "Color-based tracking of heads and other mobile objects at video frame rates," In Proc. Of IEEE CVPR, pp. 21-27, 1997.
- [8] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfunder: Real-Time Tracking of the Human Body," SPIE Vol. 2615, 1995.
- [9] M. Hunke and A. Waibel, "Face locating and tracking for human-computer interaction," Proc. Of the 28<sup>th</sup> Asilomar Conf. On Signals, Sys. and Comp., pp. 1277-1281, 1994.
- [10] K. Sobottka and I. Pitas, "Segmentation and tracking of faces in color images," Proc. Of the Second Intl. Conf. On Auto. Face and Gesture Recognition, pp. 236-241, 1996.
- [11] M. Swain and D. Ballard, "Color indexing," Intl. J. of Computer Vision, 7(1) pp. 11-32, 1991.
- [12] M. Kass, A. Witkin D.Terzopoulos, "Snakes: Active contour Models," Int. J. of Computer Vision (1) #4, pp. 321-331, 1988.
- [13] C. Vieren, F. Cabestaing, J. Postaire, "Catching moving objects with snakes for motion tracking," Pattern Recognition Letters (16) #7, pp. 679-685, 1995.
- [14] A. Pentland, B. Moghaddam, T. Starner, "View-based and Modular Eigenspaces for face recognition," CVPR'94, pp. 84-91, 1994.
- [15] M. Isard, A. Blake, "Contour tracking by stochastic propagation of conditional density," Proc. 4<sup>th</sup> European Conf. On Computer Vision, Cambridge, UK, April 1996.

## AUTHOR'S BIOGRAPHY



**B.MUTHUKUMAR** received B.E (CSE) degree from Anna University, Chennai in the year 2005. M.Tech (CSE) with Gold Medal from Dr. MGR. University in the year 2007, and Distinction with Honors. Now Pursuing Ph.D in St. Peter's University, Chennai from 2008. Having 6 Years of Teaching Experience. His areas of interest are Image Processing and Networks. He published 04 International Conference and 15 National Conference and 28 National Workshops/FDP/Seminars etc., He is a member of ISTE, CSI, IEEE, Member of IACSIT and Member of IAENG.



**Dr.S.RAVI** received DEEE, A.M.I.E (ECE), and M.E (CS) degree in the year 1989, 1991 and 1994 respectively and the Doctorate degree from Anna University, Chennai in the year 2004. He is having 16 years of teaching experience in engineering college. His areas of interests are VLSI, Embedded Systems, Image Processing and Simulation and Modeling. He has published more than 25 papers in International / National Conferences and 25 in Journals. He is a member of IEEE and ISTE and Life member of IETE and IETUK.

# Using RFID to Enhance Mobile Banking Security

Zakaria Saleh  
MIS Department, IT faculty  
Yarmouk University  
Irbid, Jordan  
zzaatreh@yu.edu.jo

Izzat Alsmadi  
CIS Department, IT faculty  
Yarmouk University  
Irbid, Jordan  
ialsmadi@yu.edu.jo

**Abstract**— Mobile banking is introducing a new generation of location-independent financial services using mobile terminals. This facilitates allowing users to make payments, check balances, transfer money between accounts and generate statements of recent transactions on their cellular phones. While providing , anywhere, anytime banking to the user, the service should be secure and security needs to be implemented at various levels, starting from the SIM card security, mobile software security, and secure customer access to banking services. Banks rely on users having their mobile phones with them all the time. Hence, as a mean for security measures, banks can send alerts, anytime, in order to provide an enhanced security and services. This paper analyzes the security issues in Mobile Banking, and proposes an improved security to the mobile banking services using RFID.

**Key words:** *Mobile banking, security, RFID, Wireless communication, Pervasive Computing, smart cards, and contactless payment, wireless security, and e-commerce.*

## I. INTRODUCTION

Mobile banking is set to reform the way people manage their money, and while Internet banking brought banks to the desktop, the Mobile banking is bringing it right into users' pockets. However, in an age of uncontrolled cyber crime, security is the primary concern. The remarkable increase in cellular phone usage has been followed by an increase in mobile fraud. Many users are concerned about the security aspect when carrying out financial transactions over the mobile network.

Mobile is often the only means of access available for millions of users in many countries. A report published by IMS [62] on Mobile Applications and Services indicates that mobile penetration in many developing markets is far higher than that of banking or fixed line infrastructure. However, lack of security is seen as the biggest deterrent to the

widespread adoption of mobile financial services. KPMG LLP examined trends in the use of mobile technology of more than 4,000 people in 19 countries worldwide, where the 91 % respondents said they had never tried banking through a mobile device, and 48% (those respondents who have not conducted banking through a mobile device) cited security and privacy as the primary reason. This research will investigate the current security within mobile banking while focusing on users' authentication, and propose a model that will further enhance access security using RFID.

### *What is mobile banking?*

The Mobile Banking environment requires both a Bank and a Mobile Network Operator (MNO) to deliver a Transactional or informational banking service to a consumer through the mobile phone. The implementation of wireless communication technologies may result in more complicated information security problems [23]. In developing countries, the role of the mobile phone is more extensive than in developed countries, as it helps bridge the digital divide. Even with initiatives like the One Laptop per Child (OLPC), the mobile penetration in many developing markets is far higher than that of banking or fixed line infrastructure [62]. People carry their mobile phones at all times, and services beyond voice communication are expected by users all over the globe. Users desire the same kind of services they get through an Internet-connected PC to be available through their mobile phone.

Mobile banking allows users to perform everyday banking functions using the mobile phone. All the major banks offer some type of mobile service for bill payment, funds transfers, checking balances, and receiving alerts [19]. Financial institution use mobile banking in one of different modes:

- **Mobile Text Banking:** In their simplest form, mobile banking services enable users to retrieve information

about bank accounts from a mobile phone using Short Message Service (SMS).

- Mobile Web/Client Banking: Using a mobile phone's data connection, this service provides users with an interface and a login with password feature.

#### Mobile Text Banking

SMS Based applications may be the simplest form of mobile banking implementation [18]. The solution is not intuitive and has no aesthetic value but is as simple as sending an SMS. SMS is used primarily as an informational banking tool as opposed to transactional banking. However, SMS can provide a pro-active functionality to send brief text messages to customers ensuring that the relevant information is provided to the user at the "right" place, at the "right" time [21]. The reason being that transactional banking requires certain levels of security, and while SMS is encrypted using the standard GSM encryption across the air, the SMS message is store in plaintext format, and the current SMS banking design has neglected the fact that some employees working for the cellular service provider can have access to the transmitted message at the service stations. Therefore using plaintext SMS message to send security details is not sufficiently secure [20]

#### Mobile Web/Client Banking

Mobile Web/Client Banking is a browser-based application, where users would access the Internet from a mobile phone. It usually offer 24/7 real-time access to users accounts right from a Web-enabled cell phone, allowing users to access account information, pay bills, transfer funds, or find a in some cases nearby ATM or Branch from the handheld mobile device[24]. The service requires no special software. However, For Mobile Web/Client Banking, the phone would have to support web browsing [22], which usually requires a "data" support plan as part of the mobile service.

The Radio Frequency Identification (RFID) system at the very simplest level, Radio Frequency Identification (RFID) system consists of a tag (or transponder) and reader (or interrogator) with an antenna. Tags can be passive with no power source or active. The technology allows for the transmission of a serial number wirelessly, using radio waves. A typical RFID transponder (tag) which can be passive (no battery) or active (with battery) consists of an antenna and an integrated circuit chip which is capable of storing an identification number and other information [16]. The reader sends out electromagnetic waves. The tag antenna is tuned to receive these waves. A passive RFID tag draws power from the field created by the reader and uses it to power the microchip's circuits. The chip then modulates the waves that the tag sends back to the reader, which converts the new waves into digital data. RFID systems use many

different frequencies, but generally the most common are low-frequency (around 125 KHz), high-frequency (13.56 MHz) and ultra-high-frequency or UHF (860-960 MHz). Microwave (2.45 GHz). The RFID operating frequencies and associated characteristics are illustrated in table 1[17].

TABLE I: RFID OPERATING FREQUENCIES AND ASSOCIATED CHARACTERISTICS.

Band	Low frequency	High frequency	Ultra high frequency	Microwave
Frequency	30–300kHz	3–30MHz	300 MHz–3GHz	2–30 GHz
Typical RFID Frequencies	125–134 kHz	13.56 MHz	433 MHz or 865 – 956MHz 2.45 GHz	2.45 GHz
Approximate read range	less than 0.5 meter	Up to 1.5 meters	433 MHz = up to 100 meters 865-956 MHz = 0.5 to 5 meters	Up to 10m
Typical data transfer rate	less than 1 Kbit/s	About 25 Kbit/s	433–956 = 30 Kbit/s 2.45=100 Kbit/s	Up to 100 Kbit/s
Typical use	Animal ID Car immobilizer	Smart Labels Contact-less travel cards	Specialist animal tracking Logistics	Moving vehicle toll

#### A smart phone with RFID tag for ATM communication: Experiments and Analysis; RFID enabled cell phones

A paper published in RFID journal in 2004 [33] predicted that within 5 years, 50% of cell phones will include RFID chips to use Near Field Communication (NFC), a two-way technology. The service was supposed to automatically connect cell phones with services in a similar fashion that occurs between airplanes and air traffic controllers on earth. NFC technology uses short-range RFID transmissions that provide easy and secure communications between various devices [33]. The important element in this proposal is the automatic peer to peer communication between RFID equipments without user involvement. The cell phone can be connected to RFID enabled applications such as websites, ATMs, restaurant outlets, GPS, etc. Files or video transfer is also possible similar to the current Bluetooth technology. In order to make this work, an NFC chip embedded in a phone can act as an RFID reader when the phone is on and a passive smart label or RFID tag when the phone is off.

There are two main ways to integrate RFID with a wireless smartphone: "A smartphone with RFID tags" and "a smartphone with an RFID reader" [34]. The first one is a typical cell phone that has embedded or attached an RFID chip with some identification information programmed on it. Its antenna is also equipped with RF antenna to be able to communicate with the RFID readers when they are within



the range. The RFID tag information is sent to the reader and the reader can write information back to the phone.

On the other hand, the second type contains an RFID reader that can collect data from various RFID tags with also an RF antenna.

However, the technology is not going very smooth. The limited UHF bandwidth and dense reader problems are still major issues to adoption

#### *NFC and ISO 14443 13.56 standard for NFC and RFID enabled phones*

Near Field Communication (NFC) is a standards-based, short-range wireless connectivity technology that enables simple and safe two-way interactions among electronic devices [61]. An ISO standard (14443) is proposed for NFC RFID enabled phones operating at 13.56 MHz in close proximity with a reader antenna. 14443 has certain features that make it particularly well-suited to applications involving sensitive information such as contactless credit cards as data transmitted is encrypted and the transmission is very short. Physical contact between the reader and the transponder is not necessary. Even a line of sight is not required. A tag may be attached to a package in the form of a smart label, worn on a person hand, attached to a ring of keys or carried in a purse along with conventional credit cards.

Some of the sought goals from using NFC RFID enabled phones are: Making payments using contactless card readers, reading account or status information from any equipment that has RFID such as stores items, discounts from smart posters or smart billboards, etc, store tickets to access transportation gates, parking garages or get into events, and many others.

## II. LITERATURE REVIEW

Recently, there are many examples for RFID enabled applications. For example, Objecs company (iwww.objecs.com) has developed three, cell-phone readable tablets suitable for gravestones that once touched can read information about the deceased. In 2005, Wal-Mart announced its decision to require its suppliers to be ready to track goods using RFID tags. Other fields of applications for RFIDs are: Transport and logistics: toll management, tracking of goods, security and access control: tracking people (students etc.), controlling access to restricted areas, supply chain management: item tagging, theft-prevention, medical and pharmaceutical applications: identification and location of staff and patients, asset tracking, counterfeit protection for drugs, manufacturing and processing: streamlining assembly line processes, agriculture: tracking of animals, quality control, public sector, sports and shopping [38]. There are some other applications that are expected to be used with RFID enabled smartphones. Examples of such applications include: web information retrieval, data transmission, automated messaging, voice services, device

integration, presence indication, and mobile payments and money transactions.

The focus on this literature review will be on FRID applications in cell phones and more particularly for banking applications. A smartphone with an RFID reader can be placed on a tag located on an equipment and use the wireless network to browse through the Internet [35]. Similar to wireless sensors, RFID enables phones can collect data at real time for many applications such as automatic material, items, weather status tracking, etc.

Currently, there are many phone companies such as Nokia, Motorola, Apple, Minec who are designing or developing RFID enabled phones [35, 36, 37]. In 2004, Nokia introduced its first RFID enabled phone 5140. Figure shows the user interface for Nokia 3220 that is also RFID enabled.

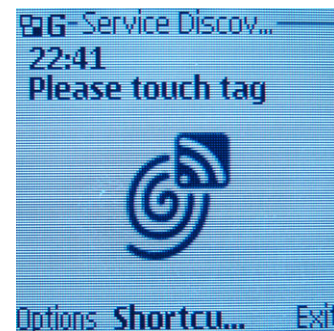


Figure 1. Cell phone screen with RFID tag feature

Mobile payment with RFID enabled phones is already available in some regions of the world. For example, in Japan and Germany, train users can pay their tickets using their enabled phones. Similar approaches are applied for airline check-in services. In France, Carrefour embraces RFID payments by card and phone.

In the following paragraphs, we will mention some papers that discussed using wireless phones in the security of mobile banking which is the focus of this subject. Some papers discussed mobile banking security, evaluations and metrics in general and examples of threats. [42, 44, 49, 50, 51, 53, 54, 56, 57]. Narendiran et al discussed using PKI security framework for mobile banking [40]. Shahreza discussed using stenography for improving mobile banking security [41]. Hossain et al [43] discussed enhancing security of SMS for financial and other services [43]. Manvi et al, Itani et al, and Krol et al proposed using J2EE and J2ME for enhancing mobile banking security [45, 47, 58]. Hwu et al proposed an encrypted identity mechanism for financial mobile security [46]. Ghotra et al proposed using Secure Display Devices (SDD) with phones for secure financial transactions [48]. Zhu et al and Rice et al proposed a framework for secure mobile payments based on cryptography [52, 55]. Henkel et al discussed the idea of



secure remote cash deposit [59]. Finally, in a similar goal to this paper, Arabo proposed utilizing phones for securing ATM transactions [60]

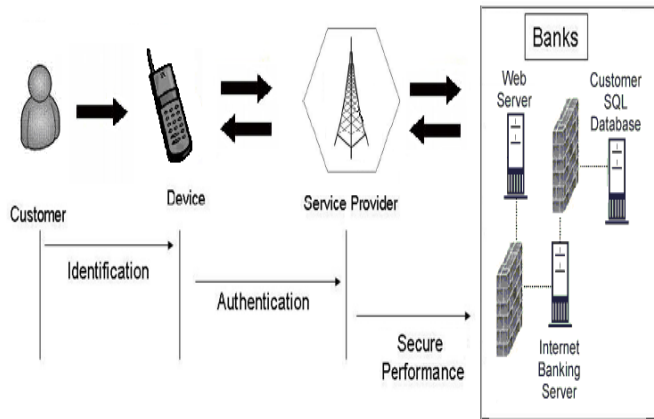


Figure 2: Mobile Banking Security System

### III. THE PROPOSED SOLUTION FRAMEWORK

#### A. Mobile Banking Security System

Figure 2 shows a typical mobile banking system using cell phones. In mobile banking as with online and traditional banking methods, security is a primary concern. Banks announce that all standard “Distance” Banking security features are applied at login including multifactor authentication by soliciting multiple answers to challenge questions. However, this may be considered strong authentication but, unless the process also retrieves 'something you have' or 'something you are', it should not be considered multi-factor. Nevertheless, Data security between the customer browser and the Web server is handled through Secure Sockets Layer (SSL) security protocol. SSL protects data in three key ways: 1) Authentication to ensure that a user is communicating with the correct server; 2) Encryption to make transferred data unreadable to anyone except the intended recipient; 3) Data integrity and verify that the information sent by users was not altered during the transfer (usually If any tampering has occurred, the connection is dropped) [6]. There are no bouts that banks have taken every precaution necessary to be sure that information is transmitted safely and securely. The security of mobile banking application is addressed at three levels (see Figure 2). The first concern is the security of customer information as it is sent from the customer's mobile phone to the Web server. The second area concerns the security of the environment in which remote access to the banking server and customer information database reside. Finally, security measures are in place to prevent unauthorized users from attempting to log into the online banking section of the Web site.

Mobile Banking gives users instant connectivity to their accounts anytime, anywhere using the browser on their mobile device, allowing users to access account details, history and check account balances, which increase convenience for the consumer, while reducing banking costs. Value-added services are the key for long-term survival online banking. However, given the uncertain nature of the transmission environment, there are security shortfalls in the present mobile banking implementations such as security problems with GSM network, SMS/GPRS protocols and security problems with current banks mobile banking solutions [63].

Services have security and privacy barriers that causes resistance and slows down the adoption, a recent study shows that 91 % of the respondents said they had never tried banking through a mobile device, and 48% of those who have not conducted banking through a mobile device indicated that security and privacy are the primary reason . A lot still prefer traditional telephone banking or ATMs and service terminals [1]. Thus, bank managers could enhance adoption of mobile banking services by concentrating their marketing efforts on factors under those barriers.

#### B. Proposed Framework Modification

Banks providing mobile services need to work on reducing security risks and improving customers' trust. Therefore, in an attempt to help banks achieve a high level of trust of mobile banking, this study has developed a module that shall further tighten security of mobile banking, and reduce the associated risk (see Figure 3), by adding a Radio-Frequency Identification (RFID) reader to the mobile banking system, on the end user's mobile phone.

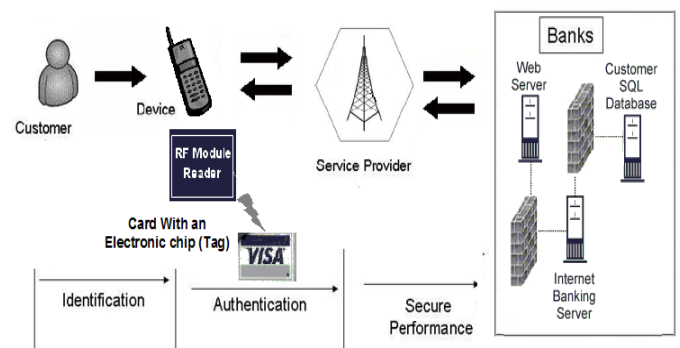


Figure 3: Proposed Module to Increase Mobile Security

- **Proposed hardware changes: Cell phones with RFID tags**

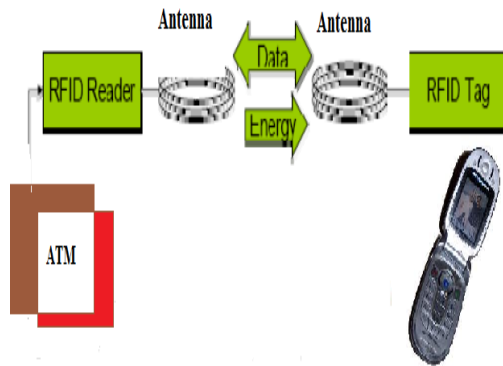


Figure 4. Connecting cell phone with ATMs through RFID

RFID tags, that are composed of an antenna connected to an electronic chip. Figure 4 shows a simple design to connect cell phones with the ATM system. When an RFID tag passes through the field of the scanning antenna, it detects the activation signal from the antenna. That "wakes up" the RFID chip, and it transmits the information on its microchip to be picked up by the scanning antenna. The RFID reader transmits radio-frequency queries, tags respond by sending back information they enclose. Finally, a Mobile phone hosting a specific RFID application pilots the reader and processes the data it sends. RFID does not require a line-of-sight reader. This whole process is depicted in Figure SSS.

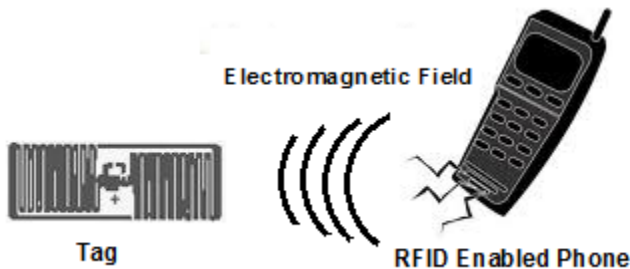


Figure 5. RFID enabled phones.

- **Proposed software changes, Programming the cell phone**

The major modification proposal for phones is hardware. Once, the phone is NFC RFID enabled, accompanied software can be included to be able to synch the phone with the RFID reader. Other expected tasks will depend whether we want the RFID tag in the phone to be active or passive, or if we want it to send and receive signals or just be a passive receiver or responder (Figure 5).

- **Programming the ATM and the banking system**

ATM user interface should be modified to include adding a new security rule for login. Figure 6 and 7 show the

proposed use login use case for ATM that include verifying customers identity with their RFID tag along with the card number and PIN.

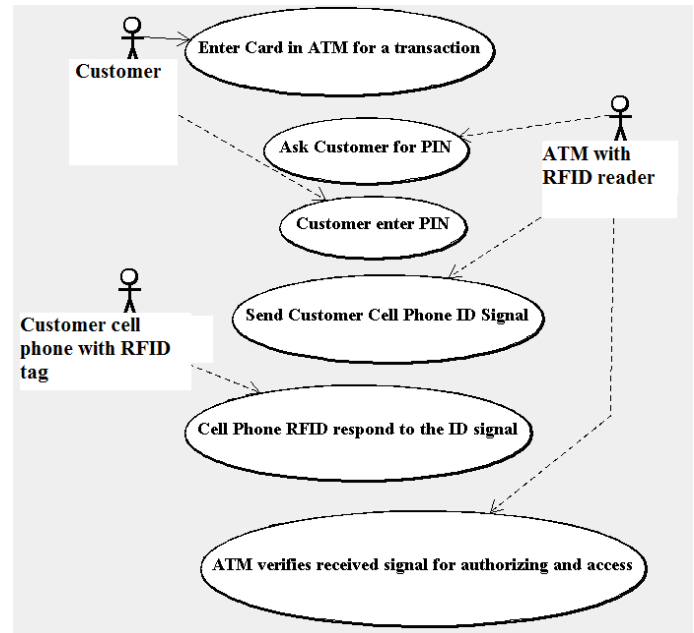


Figure 6: Use case of proposed modification on ATM access authentication

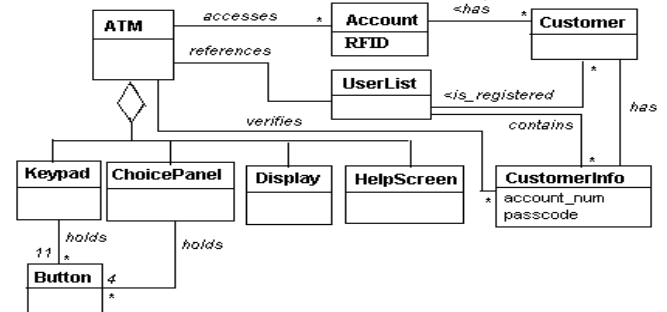


Figure 7. Typical ATM display model, with RFID attribute added to customer accounts

The banking system should be also modified to be able to deal with users RFID tags creation, cancelation, update, verification, etc. Eventually this can be incorporated with the database management system where the tag ID will be added as an attribute to users' accounts.

### C. CONCLUSION AND FUTURE WORK

In this paper, we proposed utilizing NFC RFID enabled phone for mobile banking security. This proposal is expected to solve problems with identity or credit card thefts. Users will be required to have their smart phones with them to be able to process ATM transactions. This is convenient as

users usually have their mobile phones with them all the time. Technology can help facilitating this service without breaking bank or users' privileges or security.

## REFERENCES

- [1] Berger, S. C., and Gensler, S. (2007) "Online Banking Customers: Insights from Germany". *Journal of Internet Banking and Commerce*, , vol. 12, no.1.
- [2] Betts, W. (2000). Defying denial of service attacks. *Network Magazine*, 16(5), 36-41
- [3] Greenberg, P. A. & Caswell, S. (February 1, 2001). Online banking fraud raises more security concerns. *E-Commerce Times*, , Retrieved August 14, 2003, <http://www.ecommercetimes.com/perl-story/?id=2390>
- [4] Cheung, C. and M. Lee (2000). "Trust in Internet Shopping: a proposed model and Measurement Instrument". *Proceedings of the Americas Conference on Information Systems*, pp. 681-689
- [5] Dandash, O., Le, P. D., and Srinivasan, B. (2007) "Security Analysis for Internet Banking Models". *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, July 30, 2007-Aug. 1 2007 Page(s):1141 - 1146
- [6] Freier A., Karlton P., and Kocher P. (1996). "SSL 3.0 Specification". draft-freier-ssl-version3-02.txt, Netscape Communications
- [7] Foster, A. (2002) "Federal Officials Issue Alert on Security of College Networks." *Chronicle of Higher Education*, July 5, 2002, A32.
- [8] Read, B. (2002). "Delaware Student Allegedly Changed Her Grades Online." *Chronicle of Higher Education*, August 2, 2002, A29.
- [9] Saleh, Z. I. (2003). "An Examination Of The Internet Security And Its Impact On Trust And Adoption Of Online Banking ", Unpublished PhD Dissertation, Capella University
- [10] Sarma, S. "Integrating RFID" *Queue*, Volume 2 Issue 7, ACM Press, 2004.
- [11] Stewart, D. Pavlou and S. Ward (2001). "Media Influences on Marketing Communications." In *Media Effects: Advances in Theory and Research*, J. B. a. D. Zillmann (Ed.), Erlbaum, Hillsdale, N. J.
- [12] Koufaris, M. and Hampton-Sosa, W. (2005). "The Effect of Web Site Perceptions on Initial Trust in the Owner Company" *International Journal of Electronic Commerce* Vol 10, No 1, Pages 55-81
- [13] Laukkanen, P., Sinkkonen, S., Laukkanen, T., and Kivijärvi, M. (2007). "Consumer Resistance and Intention to Use Internet Banking Services. EBRF 2007 conference, 25-27 September 2007. Finland
- [14] Want, R. "RFID Magic" *Queue*, Volume 2 Issue 7, ACM Press, 2004.
- [15] Woodforest (2007). "Frequently Asked Questions". Retrieved August 12, 2007 <<http://www.woodforest.com/default.aspx>>.
- [16] Galehdar, A. Thiel, D & O'Keefe S (2007). "Antenna Efficiency Calculations for Electrically Small, RFID Antennas" *IEEE Antennas and Wireless Propagation Letters*, VOL. 6, 156-159.
- [17] IET (2006). *Radio Frequency Identification Device Technology (RFID) Factfile*. The Institution of Electrical Engineers. <http://www.iee.org/Policy/sectorpanels/control/rfid.cfm>
- [18] Mallat, N, Rossi, M, & Tuunainen, V. (2004). "Mobile Banking Services". *Communications of The ACM*, Vol. 47, No. 5. 42-46
- [19] Adler, J. (2009) "Is Mobile Banking Getting Connected?". *DIGITAL TRANSACTIONS.NET*, VOL 6 No. 6. P 28-33
- [20] Chong M (2006). "Security of Mobile Banking: Secure SMS Banking ". *Data Network Architectures Group*. University of Cape Town, South Africa
- [21] Rajnish Tiwari, R. Buse, S. & Herstatt C. (2006) "Mobile Banking As Business Strategy: Impact Of Mobile Technologies On Customer Behaviour And Its Implications For Banks". *Portland International Conference on Management of Engineering and Technology (PICMET) 2006*, 8-13 July 2006, Istanbul, Turkey.
- [22] Kuwayama, J. (2008) "New Mobile Banking Products Present Opportunities And Challenges". Printed in Wisconsin Community Banking News June 2008.
- [23] Jin Nie Xianling Hu (2008). "Mobile Banking Information Security and Protection Methods". *Computer Science and Software Engineering*, 2008 International Conference on, 12-14 Dec. 2008, 587 - 590
- [24] Deb M. (August 2009). "Keep Your Finances Literally at the Tip of Your Fingers". *Bank of America Mobile Banking*, REVIEW - Retrieved on March 2010 from [www.appshouter.com/iphone-app-review---bank-of-america-mobile-banking/](http://www.appshouter.com/iphone-app-review---bank-of-america-mobile-banking/)
- [25] Mohammed A Qadeer, Nadeem Akhtar, Shalini Govil, Anuja Varshney, A Novel Scheme for Mobile Payment using RFID-enabled Smart SIMcard, 2009 International Conference on Future Computer and Communication
- [26] Jiahao Wang1, 2, Edward C. Wong2, Terry Ye3, PGMAP: A Privacy Guaranteed Mutual Authentication Protocol Conforming to EPC Class 1 Gen 2 Standards, *IEEE International Conference on e-Business Engineering*, 2008.
- [27] Jiahao Wang1, 3, Terry Ye2, Edward C. Wong3, Privacy Guaranteed Mutual Authentication on EPCglobal Class 1 Gen 2 Scheme, *The 9th International Conference for Young Computer Scientists*, 2008.
- [28] Ching-Nung Yang, Jie-Ru Chen, Chih-Yang Chiu, Gen-Chin Wu, and Chih-Cheng Wu, Enhancing Privacy and Security in RFID-Enabled Banknotes, 2009 IEEE International Symposium on Parallel and Distributed Processing with Applications.
- [29] D. Malocha, N. Kozlovski, B. Santos, J. Pavlina, M. A. Belkerdid and TJ Mears, II, ULTRA WIDE BAND SURFACE ACOUSTIC WAVE (SAW) RF ID TAG AND SENSOR, *Military Communications Conference, 2009. MILCOM 2009. IEEE*.
- [30] Anand Oka and Lutz Lampe, Distributed Scalable Multi-Target Tracking with a Wireless Sensor Network, *IEEE Communications Society*, 2009.
- [31] Xu Guangxian, The Research and Application of RFID Technologies in Highway's Electronic Toll Collection System, *Wireless Communications, Networking and Mobile Computing*, 2008.
- [32] Mohamed Gamal El Din, Bernd Geck, and H. Eul, Adaptive Matching for Efficiency Enhancement of GAN Class-F Power Amplifiers, *IEEE MTT-S International Microwave Workshop on Wireless Sensing*, 2009.
- [33] Claire Swedberg, Developing RFID-Enabled Phones, *RFID Journal*, July 9th 2004.
- [34] Dora Karali, Integration of RFID and Cellular Technologies1, Technical report/ white paper UCLA-WINMEC-2004-205-RFID-M2M.
- [35] Nokia's RFID Kit, <http://www.nokia.com/cda1?id=55739>.
- [36] RFID Journal, Nokia Unveils RFID Phone Reader, March 17, 2004, Gerhard Romen
- [37] Minec Web Site: <http://www.minec.com/>
- [38] Christoph Seidler, RFID Opportunities for mobile telecommunication services, *ITU-T Lighthouse Technical Paper*, 2005.
- [39] Elham Ramezani, Mobile Payment, 2008. < <http://webuser.hs-furtwangen.de/~heindl/ebte-08-ss-mobile-payment-Ramezani.pdf>>.
- [40] C. Narendiran1 S. Albert Rabara, N. Rajendran, PUBLIC KEY INFRASTRUCTURE FOR MOBILE BANKING SECURITY, *Proceedings of the World Wireless Congress, WWC'2008*
- [41] Mohammad Shirali-Shahreza, Improving Mobile Banking Security Using Steganography, *International Conference on Information Technology (ITNG'07)*.
- [42] Jin ,Nie, Xianling, Hu, Mobile Banking Information Security and Protection Methods, 2008 International Conference on Computer Science and Software Engineering
- [43] Md. Asif Hossain1, Sarwar Jahan, M. M. Hussain, M.R. Amin, S. H. Shah Newaz, A Proposal for Enhancing The Security System of Short Message Service in GSM. 235-240, *ASIS ISBN: 978-1-4244-2585-3*, 2008.
- [44] C.Narendiran, S.Albert Rabara, N.Rajendran, Performance Evaluation on End-to-End Security Architecture for Mobile Banking System, *Wireless Days*, 2008. WD '08. 1st IFIP
- [45] S. S. Manvi, L. B. Bhajantri, Vijayakumar.M.A, Secure Mobile Payment System in Wireless Environment, 2009 International Conference on Future Computer and Communication
- [46] Jing-Shyang Hwu, Rong-Jaye Chen, and Yi-Bing Lin, An Efficient Identity-based Cryptosystem for End-to-end Mobile Security, *IEEE*

TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 5,  
NO. 9, SEPTEMBER 2006

- [47] Wassim Itani and Ayman I. Kayssi, J2ME End-to-End Security for M-Commerce, Wireless Communications and Networking - WCNC 2003
- [48] Sandeep Singh Ghotra, Baldev Kumar Mandhan, Sam Shang Chun Wei, Yi Song, Chris Steketee, Secure Display and Secure Transactions Using a Handset, Sixth International Conference on the Management of Mobile Business (ICMB 2007)
- [49] Mahesh .K. harma , Dr. Ritvik Dubey, Prospects of technological advancements in banking sector using Mobile Banking and position of India, 2009 International Association of Computer Science and Information Technology
- [50] Jongwan Kim, Chong-Sun Hwang, Applying the Analytic Hierarchy Process to the Evaluation of Customer-Oriented Success Factors in Mobile Commerce, Services Systems and Services Management, 2005. Proceedings of ICSSSM '05.
- [51] Toshinori Sato, and Itsujiro Arita, In Search of Efficient Reliable Processor Design, Proceedings of the 2001 International Conference on Parallel Processing,
- [52] Y. Zhu and J. E. Rice, A Lightweight Architecture for Secure Two-Party Mobile Payment, 2009 International Conference on Computational Science and Engineering.
- [53] Matthew Freeland, Hasnah Mat-Amin, Khemanut Teangtrong, Wichan Wannalertsri, Uraiporn Wattanakasemsakul, Pervasive Computing: Business Opportunity and Challenges, Management of Engineering and Technology, 2001. PICMET '01.
- [54] Zhenhua Liu and Qingfei Min, and Shaobo Ji, An Empirical Study on Mobile Banking Adoption: The Role of Trust, 2009 Second International Symposium on Electronic Commerce and Security
- [55] J. E. Rice and Y. Zhu, A Proposed Architecture for Secure Two-Party Mobile Payment, IEEE PacRim09
- [56] Toshinori Sato'y, and Itsujiro Arital, Evaluating Low-Cost Fault-Tolerance Mechanism for Microprocessors on Multimedia Applications, Proceedings of the 2001 Pacific Rim International Symposium on Dependable Computing.
- [57] Shan chu, and Lu yao-bin. The effect of online-to-mobile trust transfer and previous satisfaction on the foundation of mobile banking initial trust, 2009 Eighth International Conference on Mobile Business.
- [58] Przemysław Krol, Przemysław Nowak, and Bartosz Sakowicz, Mobile Banking Services Based On J2ME/J2EE, CADSM'2007.
- [59] Joseph Henkel, and Justin Zhan. Remote Deposit Capture in the Consumer's Hands, IEEE 2010.
- [60] Abdullahi Arabo, Secure Cash Withdrawal through Mobile Phone/Device, Proceedings of the International Conference on Computer and Communication Engineering 2008.
- [61] Patrick Henzen, Near Field Communication Technology and the Road Ahead, NFC Forum, 2007.
- [62] IMS (2009). "900M Users for Mobile Banking and Payment Services in 2012 - 29 May 2008". Research Published July 8, 2009.
- [63] Chikomo, K., Chong, M., Arnab, A. & Hutchison A. (2006). "Security of Mobile Banking". Technical Report CS06-05-00, Department of Computer Science, University of Cape Town.

# Parcel Management System using GPS Tracking Unit

Tahmina Khatoon<sup>1</sup>, Md. Musfique Anwar<sup>2</sup>, Nasrin Sultana Shume<sup>2</sup>, Md. Mizanur Rahman<sup>1</sup>

<sup>1</sup> Computer Science & Engineering Department, IBAIS University  
Dhaka, Bangladesh

<sup>2</sup> Computer Science & Engineering Department, Jahangirnagar University  
Dhaka, Bangladesh

E-mail: tahmina8765@gmail.com, musfique.anwar@gmail.com, shume\_sultana@yahoo.com,  
mmrcse93@yahoo.com

## Abstract

*In this paper, the proposed system overcomes the deficiencies of parcel management system for providing parcel information by using operational data extracted from a GPS tracking unit. A GPS tracking unit is a device that uses the Global Positioning System to determine the precise location of a vehicle to which it is attached and to record the position at regular intervals. The recorded location data can be transmitted to a central location database of a remote tracking server using satellite modem embedded in the unit. Tracking server also has satellite modem that receives vehicle location information. This allows the vehicle's location to be displayed against a map backdrop in real-time using customized software to authorized users of the system via website over the internet.*

**Keywords:** Parcel Management System, GPS, Tracking server, Satellite Modem.

## 1. Introduction

Now a day, several companies like UPS, DHL, FedEx etc are able to provide parcel information through online. But their updating procedure of parcel information is manual. In existing system, it is not possible for a customer to know how first his/her parcel is moving or what is the location of the parcel just in this moment. But now people want to get information as early as possible. They want to know more exact and specific information about their parcel. So the next time when they stick a product in a parcel post package for shipment across the globe, they always expect the developments that have occurred in the industry to make their packages move a little bit faster.

In this paper, we propose a way to get parcel information more specifically using GPS tracking system. GPS is a Satellite Navigation System funded, controlled and operated by the U. S. Department of Defense [1]. To track a cargo as well as a parcel, a GPS tracker will be attached with each cargo. GPS receiver calculates its position by carefully timing the signals sent by the constellation of GPS

satellites high above the Earth. Then GPS data pusher transmits cargo position, speed, velocity etc to a central location data base using a satellite modem embedded in the unit. A GPS Data pusher [2] is a device which pushes (i.e. "sends") the position of the device, at regular intervals, to a determined server that can instantly analyze the data. By this way, server database will update and customer will be able to get more exact and appropriate information. It is possible to display the location from the received position information with a map using related software which is fully compatible with Google Earth Software and accurately maps the GPS coordinates of the cargo. GPS maps are clearly displayed with comprehensive information. Thus allow to know if a vehicle is on time or late, or is doing its assigned route. It allows pinpointing the exact site of a possible robbery. GPS Worldwide Technology is easy to use, quick to install and affordably priced.

## 2. Collection of data using GPS Technology

In this section, working procedure of the proposed system is described. It starts with a short description of GPS Technology and follows the description of network structure.

### 2.1 GPS Technologies

Global Positioning System (GPS) is comprised of 24 U.S. government owned satellites that circle 12,000 miles above the earth, twice a day in precise orbits, so that several are always in view from any position [3]. The system is designed to provide worldwide positioning services with an accuracy ranging from 10 to 15 meters. Instant location information enables users to ascertain exactly where their vehicles or assets are at anytime, anywhere in the world. Due to minor timing errors and satellite orbit errors, however, more precise accuracies are unattainable with standard GPS. Atmospheric conditions can also affect GPS signals and their arrival time on Earth.

There are two GPS services in operation:

- i) Precise Position Service (PPS): This service is provided to the U.S. Department of Defense and authorized associates for security purposes.
- ii) Standard Positioning Service (SPS): This service is available at no charge to all worldwide civilian users [4].

Initially designed as a military system by the U.S. Department of Defense to improve tracking capabilities of its targets, GPS has developed into a worldwide utility with multi-use services from stand alone applications to integrated, more embedded ones. Every manmade innovation from cars and planes and ships down to cell phones and wristwatches can be outfitted with GPS technology. Over the last decade, increased numbers of emergency, business and even family vehicles carry GPS devices and systems for their various tracking needs.

## 2.2 Position Determination with GPS

All GPS tracking devices work around a few basic geometric principles. These principles are summed up in trilateration. Trilateration uses the known location of two or more points of reference and the measured distance between the subject and the reference points to give an exact location for the subject.

For example, a cargo as well as a GPS receiver stands at point B is shown in Fig. 1. Let us imagine that cargo is standing out

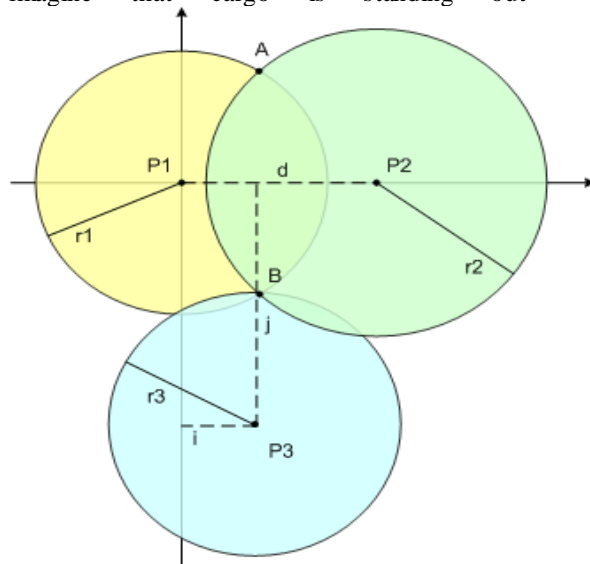


Fig. 1 Determine the cargo position with GPS receiver

in the middle of nowhere and we want to find out where it is. Our attached GPS receiver is ready for instances just like this one. P1, P2,

and P3 represent satellites in the NAVSTAR GPS constellation. These satellites are constantly transmitting microwave signals. At the same time, these satellites transmit a signal with information concerning their current location and the time that they were transmitted. The colored circles represent the path of the signal from the satellite at a specific time.

Our GPS tracking device is designed to pick up these signals and record the time at which they arrive. Because the signals from the satellites travel at approximately the speed of light, knowing the difference between when the signal was generated and when GPS receiver picks it up. Then it calculates the distance between cargo and the broadcasting satellite. This is a very simple physics equation ( $\text{speed} \times \text{time} = \text{distance}$ ). When we combine this with the known location of each of the satellites (this information is contained in the signal that GPS receiver picks up), our GPS tracking device has all the information it needs to give us an exact location for the cargo. As we use a data pusher, so it transmits cargo position information at specified intervals to a central database. This database can often be accessed through the internet, providing valuable information for our customer.

## 3. System Design and Implementation

Overall system is partitioned into two major design units.

- Cargo unit
- Tracking Server.

### 3.1 Cargo Unit

This is major part of the system and it will be installed into the cargo. It is responsible for capturing the cargo location and transmitting this location information to the remote tracking server located anywhere in the world. Cargo unit use the following modules to achieve this functionalities.

#### 3.1.1 GPS Receiver

Cargo unit uses GPS receiver to capture the current location and other information like speed, velocity etc of the cargo. Location and other information provided by GPS are not in human understandable format. This raw data needs to be processed to convert it into useful information that can be displayed on the map. CPU is required to process this raw data. The m100 smart satellite modem is used for data



transmission that offers world wide satellite connectivity and GPS capabilities for a wide variety of asset-tracking and industrial remote communications. Operating on the ORBCOMM low-earth orbit (LEO) satellite network, the m100 combines Machine-to-Machine (M2M) communication, unlimited global coverage without blockage, integrated GPS-based location awareness, and powerful on-module application development options in a single, highly integrated module.

### 3.1.2 Central Processing Unit

The raw data provided by the GPS receiver is captured by the CPU and processed to extract the required location and other information. CPU holds all the required information that is to be transmitted to remote server. It also controls data transmission module to exchange information with remote server. It actually acts as a bridge between GPS receiver and remote server. As the processing required in the Cargo unit is not computationally intensive therefore any low end microcontroller can be used as a CPU. The microcontroller selected to serve as CPU for Cargo unit is Microchip's PIC18F248. This is 8-bit microcontroller and runs at speed of 20 MHz which is enough speed for the system.

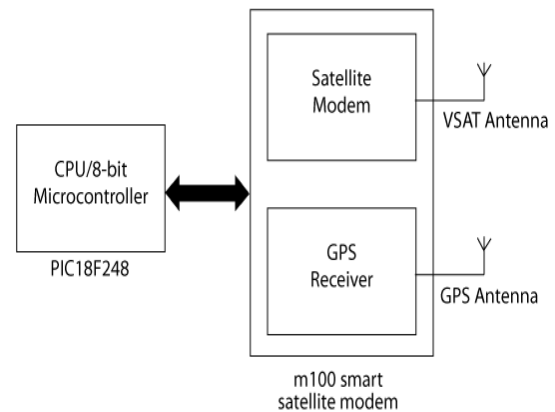
### 3.1.3 Data Transceiver

When all required information is extracted and processed, it needs to be transmitted to a remote tracking server which will be able to display this information to the end user. For real time tracking of vehicle, reliable data transmission to remote server is very important. Wireless network is required to transmit cargo information to remote server. VSAT [5] technology is selected to do this. VSAT is a way of providing better connectivity to the internet or for private satellite communications networks. Depending on bandwidth requirement (data speed and/or communications channels), VSAT systems can be relatively small (1.2 to 2.4 meter antenna) and easily installed. By linking VSAT terminals to larger hub stations (or land earth stations), a wireless network can be established.

### 3.1.4 Cargo Unit Design

Cargo unit is designed using m100 smart satellite modem and microcontroller

PIC18F248 manufactured by Microchip. Fig. 2 shows the block diagram of cargo unit.



**Fig. 2 Cargo unit block diagram**

GPS antenna receives signals from GPS satellites. It must face towards sky for correct computation of the current location by GPS receiver. Location data is transferred to microcontroller through serial interface. After processing of the data provided by GPS receiver, microcontroller transmits this information to remote location using embedded satellite modem. Microcontroller controls the operation of m100 smart satellite modem through serial interface. External VSAT antenna is required by the m100 smart satellite modem for reliable transmission and receiving of data.

#### 3.1.4.1 m100 satellite modem

The m100 is the next generation satellite data communication transceiver that transmits and receives data by using the ORBCOMM [6]. The m100 is a low cost ORBCOMM modem based on the Analog Devices Blackfin DSP and analog ASIC. The m100 OEM transceiver plays a vital role in transmitting and receiving the data for the customers involved in satellite-based tracking and industrial remote communication. The m100 operates over the ORBCOMM low orbit satellite network, providing unlimited global coverage with no blockage. It can significantly improve asset utilization by allowing clients to monitor, track and manage their fixed and mobile assets around the world. The three chip m100 OEM transceiver includes the Analog Devices Blackfin® DSP family, MobiApps' Analog ASIC, and a GPS RF down converter, creating a fully integrated satellite tracking device.

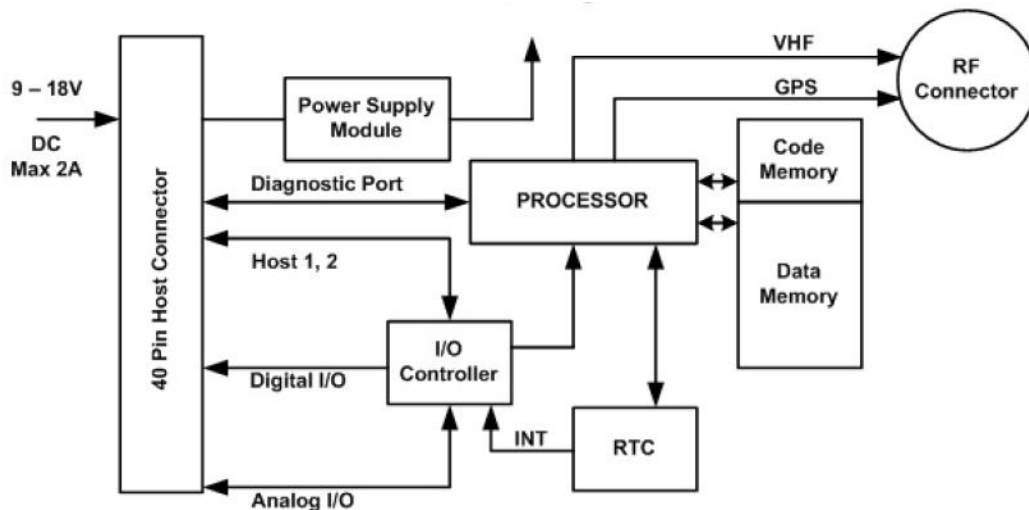


Fig. 3 Block diagram of m100 satellite modem

### 3.1.4.2 ORBCOMM working procedure

The ORBCOMM system provides global, two-way, data communication services to a wide variety of applications [6]. Subscriber communicators (SCs) pass data messages to and from Gateway Control Centers (GCC) over ORBCOMM satellites and GCCs route messages to users over the internet or dedicated delivery lines. SCs are a highly versatile communication device which communicates

directly with satellites using ORBCOMM's packet-switched protocol, and supports full transmission acknowledgement. The satellites provide SCs with system information and serve as the communication link between SCs and the ORBCOMM terrestrial network. Messaging traffic flows between the satellites and a GCC through tracking stations called Gateway Earth Stations (GESs) that connect with satellites as they pass overhead. The internet is commonly used for final transmission to the user.

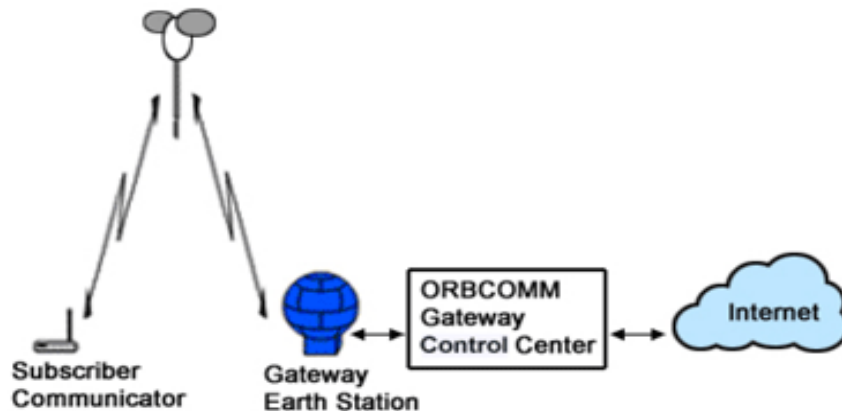


Fig. 4 ORBCOMM working procedure

### 3.1.4.3 Cargo unit software design

Microcontroller is acting as Central Processing Unit for cargo unit. All operations of the cargo unit are to be controlled by the microcontroller. Microcontroller needs instructions to operate the whole system. These instructions are

provided to microcontroller by writing the software into microcontroller's flash memory. It reads the software instruction by instruction and performs the action as required by instruction. Complete software is broken down into small modules as shown by the Fig. 5.

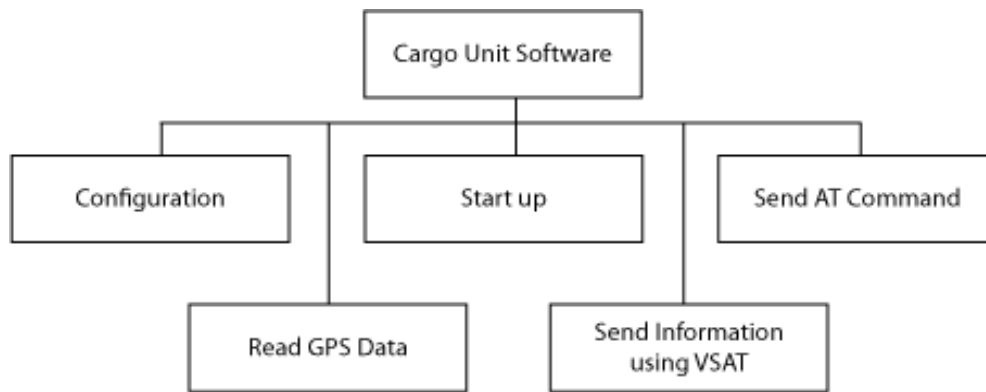


Fig. 5 Breakdown of cargo unit software

#### 3.1.4.4 Subroutine-Send AT Command

This subroutine is the basic routine which handles all the communication with m100 satellite modem. This routine accepts the string containing "AT" command input in its parameters and sends this string character by character to module [7]. GPS receiver accepts carriage return ('\r') as a command terminating character. As this character is received it sends back the response to microcontroller. Fig. 6 shows the flowchart of Subroutine-Send AT Command.

#### 3.2 Tracking Server

Tracking server maintains all information received from all Cargo units installed in different cargos into a central database. This database is accessible from internet to authorized users through a web interface. Tracking server has a m100 satellite modem attached to it that receives information from cargo units and sends those messages to the server through serial port. Tracking server saves this information into database.

### 4. Working Principle of the Entire System

When cargo moves its integrated GPS receiver send information of each position through internet to a remote Server database. The GPS receiver determines its current location, speed and heading. This data can be directly transmitted from the operating unit of the cargo to the server and this information can be stored in server database. How this position will be calculated and how the encrypted information will be send to server database has been already discussed. The main component of the server is an interface with the name "ServerReceiver". This class is able to respond to HTTP or UDP requests. To generate an

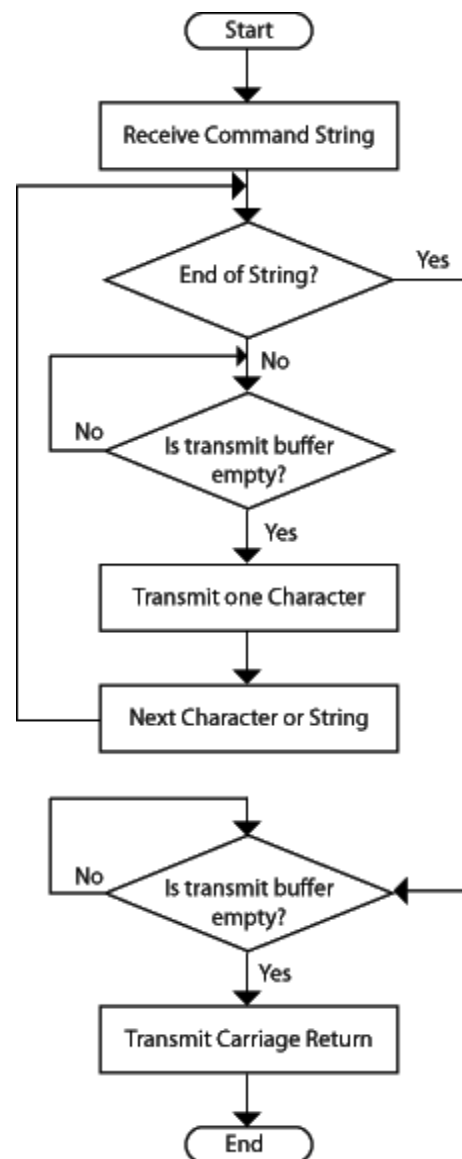


Fig. 6 Flow-chart of Subroutine-Send AT Command

answer the class has to read a so called "InputStream". To do this the class uses another class named "InputStream" as interface to access the serial port. The class "SerialPort" can

read data from the serial port. But for this it uses a device dependent routine. With this routine the data from the receiver can send to the main class. The server will be programmed in Java and it can be interpreted with every Java

Virtual Machine (JVM) that supports Java 1.1 or Personal Java. Current position can be displayed on a PC in digital maps. Fig. 7 illustrates the entire system process.

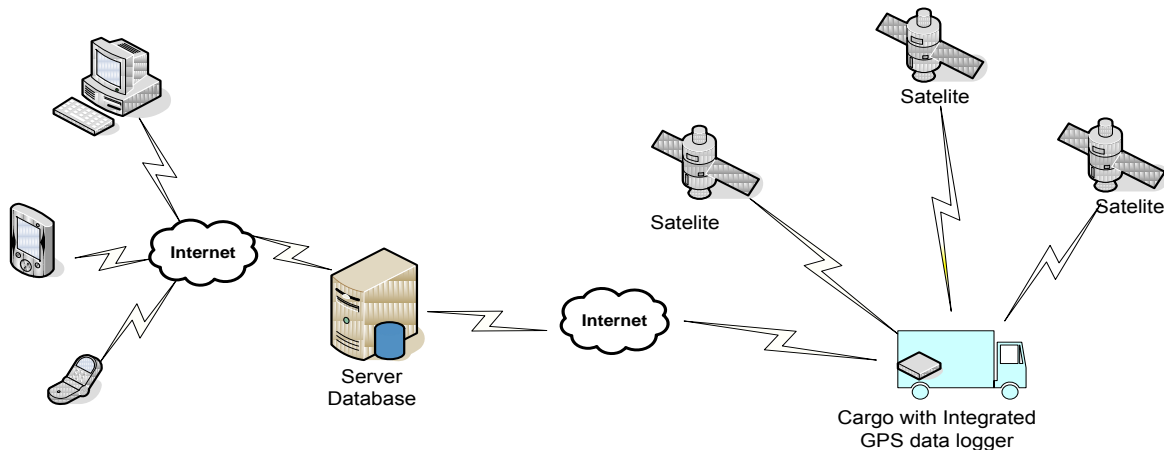


Fig 7 Interaction between cargo unit, tracking server and server

## 5. WEB Interface Design

As described in previous section, Tracking Server maintains all information in a database. The administrator of the system who is managing Parcel tracking System must have to install the required software. There must be a number of cargos installed with Cargo units and therefore server must be able to manage and distinguish information sent by all cargo units. For this purpose information must be available

to server about all cargos that are installed with cargo units. Whenever Cargo unit is installed, information about that cargo is stored in the database. Web interface supports this functionality. Since web interface will be accessible over the internet therefore access must be restricted to authorized users only. Therefore information about all users of the system must be stored in database. Fig. 8 shows the interaction between customer and server.

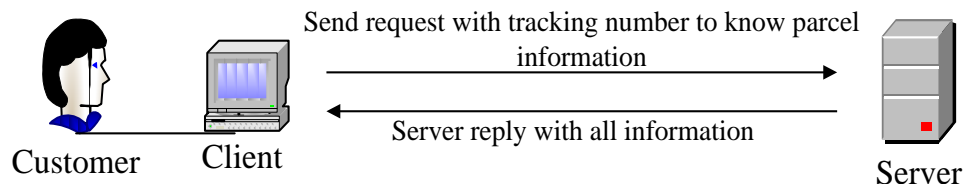


Fig. 8 Interaction between customer and server

## 6. Conclusion

Both cargo unit and tracking server side implementation has been described in this paper. Here we use VSAT technology to set up the entire wireless network instead of existing GSM/GPRS cellular network. It makes the system a bit costly then the existing one but it is reasonable if we consider its performance. This

system can also work properly anywhere in the world even where GSM network is unavailable (i.e. in the sky or under the sea). As set up of VSAT wireless network is a bit costly, so only large companies will get the maximum benefits from it. But we hope in near future the technology will be improved and it will be suitable for all types of parcel shipping companies.

## References

- [1] Interface control document, Navstar GPS Space Segment (Navigation User Interfaces), 2000.
- [2] [http://en.wikipedia.org/wiki/GPS\\_tracking](http://en.wikipedia.org/wiki/GPS_tracking).
- [3] <http://www.landairsea.com/gps-tracking-blog/>.
- [4] G. M. Djuknic and R. E. Richton, "Geolocation and Assisted GPS," Computer, vol. 34, no. 2, pp.123-125 (2001).
- [5] M. E. Jackson, C. Meertens, O. Ruud, S. Reeder, W. Gallaher, C. Rocken, "Real Time GPS Data Transmission Using VSAT Technology", submitted to GPS Solutions (May 2001).
- [6] <http://ftp1.digi.com/support/documentation/>

- m100usermanualmay07.pdf.  
[7] Muruganandham, P.R.Mukesh, "Real Time Web based Vehicle Tracking using GPS",

Journal of World Academy of Science, Engineering and Technology Issue 61, pp. 91 – 99 (2010).



**Tahmina Khatoon** completed her B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, IBAIS University, Bangladesh in 2008. She is now working

as a software developer in a renowned press media in Dhaka, Bangladesh. Her research interests include Software Engineering, Telecommunication, Image Processing and so on.



**Md. Musfique Anwar** completed his B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, Jahangirnagar University, Bangladesh in 2006. He is now a Lecturer in the

Dept. of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh. His research interests include Artificial Intelligence, Natural Language Processing, Image Processing, Pattern Recognition, Software Engineering and so on.



**Nasrin Sultana Shume** completed her B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, Jahangirnagar University, Bangladesh in 2006. She is now a Lecturer in the

Dept. of CSE, Green University of Bangladesh, Mirpur, Dhaka, Bangladesh. Her research interests include Artificial Intelligence, Neural Networks, Image Processing, Pattern Recognition, Database and so on.



**Md. Mizanur Rahman** completed his B.Sc (Engg.) in Computer Science and Engineering from Dept. of CSE, Bangladesh University of Engineering and Technology (BUET), in

2000. He is now an assistant professor in the Dept. of CSE, IBIAS University, Dhanmondi, Dhaka, Bangladesh. His research interests include Data communication, Artificial Intelligence, Software Engineering and so on.

# Generation of Mutation Operators for AOP

MAYANK SINGH\*

Research Scholar

Uttarakhand Technical University, Dehradun, India

[mayanksingh2005@gmail.com](mailto:mayanksingh2005@gmail.com)

SHAILENDRA MISHRA

Professor computer science department

KEC dwarahat, Uttarakhand, India

[skmishra1@gmail.com](mailto:skmishra1@gmail.com)

**Abstract:** Testing of aspect oriented programs is an upcoming challenge for the researchers. Mutation testing has a lot to be undertaken to explore the field of testing of AOP. It is an emerging field of research in testing of aspect oriented programming. Since the effectiveness of mutation testing depends on finding fault types and designing of mutation operators, therefore the effectiveness of testing depends upon the quality of these mutation operators. A detailed study has been done on the mutation operators for procedural and object oriented languages, but for aspect oriented language only few researchers have contributed. This paper discusses in detail about the fault types and related mutation operators for AspectJ language. It also proposes the implementation framework of mutation operators automatically.

**Keywords:** Mutation Testing, Aspect oriented testing, fault based testing

## I. INTRODUCTION

Software testing is a very crucial part of software engineering. If the testing of software is not appropriate then there is no guarantee of the quality of software product. With the help of testing process we can ensure that software realizes all the required functions as well as check the performance of software. The testing process must be done with the intention of finding bugs in the programs. In the software development life cycle, testing activities start from requirement phase and go along with all the intermediate process of development. Along with quality process, testing is the only activity which is carried out even after the development. Testing is necessary to develop any software system. Generally testing cost half of the total cost of software development and maintenance. There are two types of software testing: first structural testing and second functional testing [2,26,34,46].

Mutation testing or fault based testing is an example of structural testing for assessing the quality of software. In mutation testing we inject the fault in the original program and test same as the original program and compare the result of both programs to check the quality of program. These faulty programs are called mutants. For example, suppose a program is written to add two numbers i.e.  $c = a + b$ , for this program mutants are  $c' = a - b$ ,  $c' = a * b$ ,  $c' = a / b$ . If output of both programs are not same on the same test cases i.e.  $c \neq c'$ , then the mutant is killed. If all the mutants are killed, functionality of the program is good and test data is adequate. On the other

hand, if the outputs of both programs are same that means mutants are alive. Mutants may be alive because of the test data is unable to distinguish the mutants or equivalent mutants.

Every software engineer wants to look for ways to improve the modularity of software. AOP is a new methodology that provides separation of crosscutting concerns (design and requirement elements that affect multiple modules) through the modularization. With the help of AOP, we implement crosscutting concerns in aspects instead of them in the core modules. Aspect weaver composes the final system by combining core and crosscutting modules through a process called weaving [6,39,47]. Aspect Oriented Programming builds on the top of existing methodologies such as object oriented programming and procedural programming. AOP has some myths like AOP doesn't solve any new problem, AOP promotes sloppy design, it breaks the encapsulation, and all the testing techniques cannot be applied on AOP.

AspectJ is an aspect oriented language which is an extension to the Java programming language. AspectJ is easy to understand for Java programmers because it uses Java as base language and provides some benefits to the language. An AspectJ compiler produces class files that conform to the Java byte code specification, any Java virtual machine can execute these files [6,7,8,9]. With the help of AspectJ language we can implement Join Points- predictable execution point, Pointcut- to select the join points, advice- an code execution point, Intertype declaration - to add attributes and methods to previously established classes and aspect - an analogy to encapsulate all the above points into Java class [1,3,7].

Testing is a big issue in aspect oriented programming. In this paper we attempt to resolve this issue with the help of mutation testing techniques. Here we classify the possible faults which can occur in AOP. Further we will design corresponding mutation operators to resolve the testing issues with aspect oriented programs. We also propose the framework for implementation of mutation operators.

The rest of the paper is organized as follows: Section 2 describes in detail the related work for fault based testing. Section 3 describes the classification of faults. Mutation operators have been described in section 4. Section 5



describes the implementation architecture. Conclusion and future work is discussed in section 6.

## II. RELATED WORK

Ferrari et al. identified some fault types for aspect oriented programs that are extend in this paper. They have also identified a set of mutation operators related to the faults. They define the operators for aspectJ language and propose generalize mutation operators for other aspect oriented languages [5].

Yves presents a tool named AjMutator for mutation analysis of point-cut descriptor. In this paper they implement pointcut related mutation operators which have been identified in the previous research. This tool leverages the static analysis by the compiler to detect the equivalent mutants automatically [21].

Romain Delamare proposes a test driven approach for the development and validation of pointcut descriptor. They designed a tool named Advice Tracer which is used to specify the expected joinpoints. To validate the process, they also develop a tool that systematically injects faults into pointcut descriptors [50].

Alexander identifies key issues related to the systematic testing of aspect oriented programs. They develop a candidate fault model for aspect oriented programs and derive testing criteria from the candidate fault model [12,13,47].

Prasanth proposes a framework which automatically finds the strength of pointcuts and determines the different versions of the expression to choose the correct strength by the developer. Based on similarity measure, the framework automatically selects and rank different versions of pointcut expression [10,23].

Xie and Zhao developed a framework, named Aspectra, which automatically generate the test inputs for testing aspectual behavior. Aspectra defines and measures aspectual branch coverage [49].

Wedyan and Ghosh present a tool for measuring joinpoint coverage from per advice and per class. This tool is based on AspectJ and Java bytecode. This tool implements the framework which is given by Xie and Zhao [11].

Mortensen and T. Alexander use the static analysis of an aspect with in a system to choose the white box criteria for an aspect such as statement coverage, context coverage and def-use coverage. They also provide a set of mutation operators related to pointcut and aspect precedence [12].

## III. CLASSIFICATION OF FAULTS

To distinguish the programs from its mutants, we need effective test cases to find faults in the program. Like other

testing techniques, the efficiency of mutation testing depends on finding faults. Any mutation system can have these faults and this paper attempts to identify almost all the faults from such mutation system which is designed to represent these faults. These faults are implemented in the form of mutation operators. Effectiveness of mutation testing is depends on the quality of mutation operators. Mutation testing is not new but with respect to AOP it is new. AOP have many new features such as pointcut, joinpoint, advice, inter type declaration, aspect and weaving. These new features of AOP introduce the potential of new faults. Previously stated faults about these features are not sufficient, so we have to identify new faults to complete the qualitative mutation testing.

Some of Java related faults can be used in finding the faults of aspect oriented programs because AspectJ program uses Java language for the base program. There are only two researchers named Baekken and Ferrari, who identified fault types and related set of mutation operators [5,24]. All of these fault types focus on the characteristics and structure of AspectJ language. This paper introduces a new set of fault types and mutation operators with the inclusion of all previously stated fault types and mutation operators. Previously stated mutation operators do not handle several fault types and all AOP features.

Faults can be classified on the basis of our exhaustive survey on testing aspect oriented programming [4,5,7,10,11,12,13,14,17,21,24]. Our analysis is based on fault models, bug patterns, pointcut descriptor, and fault classifications [35,38,45,47,49]. In AspectJ language, we can find faults in a program with woven aspect i.e. a fault may exist in the base program that is not affected by the woven aspect or fault can exist in the aspect code [13,51]. Faults can be classified on the basis of pointcut, advice, java program, and intertype declaration and weaving. We have identified some new faults which are given below:

- Visibility of joinpoints or pointcut selection fault because pointcut expression selects joinpoint as it was supposed to or was not selected and neither supposed to or both ignored and unintended joinpoints or selects only ignored joinpoints or selects only unintended joinpoints.
- Faults during combining individual pointcut in conditional operators
- Incorrect use of methods, type in pointcut expression
- Use of wrong filed or constructor pattern in pointcut expression
- Use of wrong primitive pointcut descriptor
- Wrong matching based on exception throwing patterns.
- Use incorrect method name in introduction
- Inconsistent method introduction overridden
- Ripple effect production in the control flow
- Wrong changes in polymorphic calls
- Wrong changes in data dependency

- Wrong changes in control dependency
- Failure to execute the proper context
- Wrong advice precedence
- Incorrect changes in inheritance hierarchy
- Modifying variables and statements
- Inconsistency in interface on dynamic binding
- Incorrect use of advice specification
- Fault in advice code to meet its post-condition
- Wrong introduction of interface
- Addition of members at wrong places in hierarchy
- Adding wrong intertype members
- Static invariants violations occur because of weaving
- Incorrect weaving of around advice
- Failure to establish state invariants
- Incorrect or missing proceed in around advice
- Failure due to unable to find required joinpoints from base code
- Fault due to member variable declaration with parent class

#### IV. MUTATION OPERATORS FOR ASPECTJ

There are mainly four types of mutation operators available namely pointcut related operators, advice related operators, waving related operators and base Java programs related operators. Due to space, the description of these mutation operators are not given here but in the next paper we will provide the implementation details with full description of all mutation operators. On the basis of fault types, specified in the previous section, the mutation operators are as follows:

- PPCM - Incorrect change of primitive pointcut by call to execution or vice versa of methods and constructors
- PPCC - Incorrect change of primitive pointcut by Initialize to reinitialize or vice versa of constructors
- PPCF - Incorrect change of primitive pointcut by cflow to cflowbelow or vice versa
- PPCT - Incorrect change of primitive pointcut by this to target or vice versa
- PPTA - Incorrect change of primitive pointcut by target to args or vice versa
- PPTW - Incorrect change of primitive pointcut by this to within or vice versa
- PPWC - Incorrect change of primitive pointcut by cflow to within or vice versa
- PPSG - Incorrect change of primitive pointcut by set to get or vice versa
- PNUD - Use incorrect user defined pointcut name
- PPFW - Pointcut positive fault due to the use of wildcard
- PNFW - Pointcut negative fault due to the use of wildcard
- PWBI - Wrong Boolean expression for if pointcut

- PICO - Incorrect use of pointcut composition operators like change OR with AND operators or vice versa
- PCON - Use of NOT composition operator where it should not be used or vice versa
- BPCO - Incorrect use of composition operators individually like more than one args, target and this pointcut composed with AND operators. Change this composition operator with other composition operator
- BMAS - Wrong introduction of methods as abstract with synchronized, final or static
- BCAS - Incorrect introduction of constructor as abstract, static, final, volatile or transit
- BPKT - Wrong use of this keyword in base program
- BPDm - Incorrect deletion of member variable initialization
- BPDp - Incorrect declaration of member variable in parent class
- ACAB - Wrong changes in advice from before to after or vice versa
- AIDP - Incorrect deletion of proceed statement
- PMDP - Missing or deletion of pointcut parameters
- PCDP - Incorrect changes in the parameter list of pointcut descriptor
- ASPD - Deletion of aspect precedence
- BCFW - Unintended control flow execution by changing warning or error statement

#### V. PROPOSED FRAMEWORK

In this paper we are proposing a framework for implementation of these fault types and mutation operators. The proposed framework attempts to automate the test data generation process. Proposed framework takes AspectJ and Java files as an input. Parse the Aspect file to find the pointcuts and advice code. Then according to the pointcut expression, we have to identify the joinpoints from the base Java code. Compile both files with ajc compiler and add the woven code into a single Java file. After this, we have to decide the testing criteria for mutation testing. Then on the basis of fault types and mutation operators, mutants have been generated and applied according to the need.

Store all the tested data into database. On the basis of first testing data, we have decided the baseline for other iterations and execute the analysis process to check whether mutants are alive or killed. Decision about the status of mutants depends on the output data that are generated from both programs i.e. original program and mutant program. On the basis of final report we find the cost of mutation testing. This framework also finds the quality of the test cases and the mutants. The effectiveness of framework depends on the generation of fault types and mutation operators. The mutation based test case generation process is shown in figure 1.

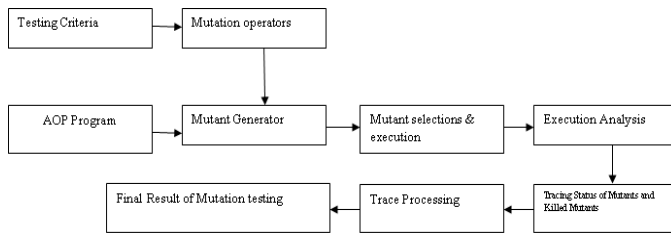


Figure 1: Mutation based test case generation

Test execution is done on the basis of generated test cases. For mutation testing at unit level, we have to identify the test objects through program slicing and then analyze the test outcome with expected outcome. Equivalent mutants are identified and killed with the help of these test objects and test execution process. Process of test execution is shown in figure 2.

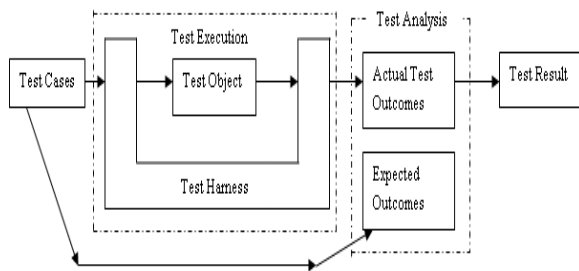


Figure 2: Test Execution Overview

## VI. CONCLUSION AND FUTURE SCOPE

This paper presents fault types and mutation operators for mutation testing related to aspect oriented programs. The operators used are based on AspectJ language which is most acceptable language for aspect oriented programming. These fault types identified from the characteristics of AspectJ language with Java language. We have also tried to identify a way to improve the efficiency and reliability of aspect oriented software by using the mutation operators based on an exhaustive list of aspect oriented faults. In this paper we proposed the implementation framework to execute the test data of mutation operators.

The extension of this study is to identify some new mutation operators and implement these operators. Our next aim is to develop an automated tool to test these mutation operators as well as generate test cases automatically. We also want to check the quality of the test data to confirm the effectiveness of aspect oriented software.

## References

- [1] AOSD-Europe Project Home Page, 2007. <http://www.aosd-europe.net/> (accessed 05/12/2010).
- [2] JBoss AOP Reference Documentation, 2007. <http://labs.jboss.com/jbossaop/docs/index.html> (accessed 05/12/2010).
- [3] The AspectJ Project, 2007. <http://www.eclipse.org/aspectj/> (accessed 02/11/2010).
- [4] R. Agrawal et al. Design of mutant operators for the C programming language. Report SERC-TR41-P, S.E. ResearchCenter, Purdue University, West Lafayette-USA, 1989.
- [5] F.C. Ferrari, J.C. Maldonado, and A. Rashid. Mutation testing for aspect oriented programs. In ICST'08, pages 52-61. IEEE Computer Society 2008.
- [6] AspectJ in Action: Practical Aspect-Oriented Programming, Manning Publications (July 1, 2003).
- [7] Mastering AspectJ: Aspect-Oriented Programming in Java, Wiley; 1st edition (March 7, 2003).
- [8] Eclipse AspectJ: Aspect-Oriented Programming with AspectJ and the Eclipse AspectJ Development Tools, Addison-Wesley Professional (December 24, 2004)
- [9] AspectJ Cookbook, O'Reilly Media; 1 edition (December 20, 2004)
- [10] Prasanth Anbalagan, Automated Testing of Pointcuts in AspectJ Programs, in OOPSLA'06, October 22–26, 2006, Portland, Oregon, USA.
- [11] Fadi Wedyan, Sudipto Ghosh, A Joinpoint Coverage Measurement Tool for Evaluating the Effectiveness of Test Inputs for AspectJ Programs, IEEE Computer Society, 2008.
- [12] Michael Mortensen and Roger T. Alexander, Adequate Testing of Aspect-Oriented Programs, Technical report CS 04-110, December 2004, Colorado State University, Fort Collins, Colorado, USA.
- [13] R. T. Alexander, J. M. Bieman, and A. A. Andrews. Towards the systematic testing of aspect-oriented programs. Report CS-04-105, Dept. of Computer Science, Colorado State University, Fort Collins-USA, 2004.
- [14] Hierons, R., Harman, M. & Danicic, S. "Using Program Slicing to Assist in the Detection of Equivalent Mutants". Journal of Software Testing, Verification and Reliability, 9(4), 233-262, 1999..
- [15] Offutt, J. & Pan, J. "Automatically Detecting Equivalent Mutants and Infeasible Paths". The Journal of Software Testing, Verification, and Reliability, Vol 7, No. 3, pages 165--192, September 1997.
- [16] King, K. N. & Offutt, J. "A Fortran Language System for Mutation-Based Software Testing". Software Practice and Experience, 21(7):686--718, July 1991.
- [17] Ma, Y. S., Offutt, J. & Kwon, Y. R. "MuJava: An Automated Class Mutation System". Journal of Software Testing, Verification and Reliability, 15(2):97-133, June 2005..
- [18] Ma, Y. S., Kwon, Y. R. & Offutt, J. "Inter-Class Mutation Operators for Java". Proceedings of the 13th International Symposium on Software Reliability Engineering, IEEE Computer Society Press, Annapolis MD, November 2002, pp. 352-363.
- [19] Ma, Y. S. & Offutt, J. "Description of Method-level Mutation Operators for Java". November 2005.
- [20] Ma, Y. S. & Offutt, J. "Description of Class-level Mutation Operators for Java". November 2005.
- [21] AjMutator: A Tool For The Mutation Analysis Of AspectJ Pointcut Descriptors: Romain Delamare, Benoit Baudry and Yves Le Traon, IEEE International Conference on Software Testing Verification and Validation Workshops 2009.
- [22] An Analysis and Survey of the Development of Mutation Testing: Yue Jia and Mark Harman, Technical Report- CREST CENTRE, KING'S COLLEGE LONDON. TR-9-06

- [23] P. Anbalagan and T. Xie. Efficient mutant generation for mutation testing of pointcuts in aspect-oriented programs. In *Mutation'2006*, pages 51–56. Kluwer, 2006.
- [24] J. S. Bækken. A fault model for pointcuts and advice in AspectJ programs. Master's thesis, School of Electrical Engineering and Computer Science, Washington State University, Pullman-USA, 2006.
- [25] E. F. Barbosa, J. C. Maldonado, and A. M. R. Vincenzi. Toward the determination of sufficient mutant operators for C. *Soft. Testing, Verif. and Reliability*, 11(2):113–136, 2001.
- [26] M. Ceccato, P. Tonella, and F. Ricca. Is AOP code easier or harder to test than OOP code? In *WTAOP'2005*, 2005.
- [27] R. Chitchyan, A. Rashid, P. Rayson, and R. Waters. Semantics-based composition for aspect-oriented requirements engineering. In *AOSD'2007*, pages 36–48. ACM Press, 2007.
- [28] D. S. Dantas, D. Walker, G. Washburn, and S. Weirich. PolyAML: A polymorphic aspect-oriented functional programming language. In *ICFP'2005*, pages 306–319. ACM Press, 2005.
- [29] M. E. Delamaro, J. C. Maldonado, and A. P. Mathur. Interface Mutation: An approach for integration testing. *IEEE Transactions on Soft. Eng.*, 27(3):228–247, 2001.
- [30] M. Gong, V. Muthusamy, and H. Jacobsen. Aspect-Oriented C tutorial, 2006. <http://research.msrg.utoronto.ca/ACC/Tutorial> (accessed 23/08/2007).
- [31] R. A. DeMillo. Hints on test data selection: Help for the practicing programmer. *IEEE Computer*, 11(4):34–43, 1978.
- [32] M. Eaddy, A. Aho, W. Hu, P. McDonald, and J. Burger. Debugging aspect-enabled programs. In *Software Composition'2007*, 2007.
- [33] R. A. DeMillo and A. J. Offutt. Constraint-based automatic test data generation. *IEEE Transactions on Soft. Eng.*, 17(9):900–910, 1991.
- [34] W. Harrison, H. Ossher, and P. L. Tarr. General composition of software artifacts. In *Software Composition'2005*, pages 194–210 (LNCS v.4089). Springer-Verlag, 2006.
- [35] O. A. L. Lemos, F. C. Ferrari, P. C. Masiero, and C. V. Lopes. Testing aspect-oriented programming pointcut descriptors. In *WTAOP'2006*, pages 33–38. ACM Press, 2006.
- [36] A. P. Mathur and W. E. Wong. An empirical comparison of data flow and mutation based test adequacy criteria. *Soft. Testing, Verif. and Reliability*, 4(1):9–31, 1994.
- [37] M. Mezini and K. Ostermann. Conquering aspects with Caesar. In *AOSD'2003*, pages 90–99. ACM Press, 2003.
- [38] S. Zhang and J. Zhao. On identifying bug patterns in aspect-oriented programs. In *COMPSAC'2007*, pages 431–438. IEEE Computer
- [39] G. Kiczales et al. Aspect-oriented programming. In *ECOOP'1997*, pages 220–242 (LNCS v.1241). Springer Verlag, 1997.
- [40] M. Stoerzer and J. Graf. Using pointcut delta analysis to support evolution of aspect-oriented software. In *ICSM'2005*, pages 653–656. IEEE Computer Society, 2005.
- [41] D. Xu and W. Xu. State-based incremental testing of aspect-oriented programs. In *AOSD'2006*, pages 180–189. ACM Press, 2006. Society, 2007.
- [42] A. J. Offutt, J. Voas, and J. Payne. Mutation operators for Ada. Report ISSE-TR-96-06, Dept. Inf. and Soft. Systems Eng., George Mason University, Fairfax-USA, 1996.
- [43] O. A. L. Lemos, A. M. R. Vincenzi, J. C. Maldonado, and P. C. Masiero. Control and data flow structural testing criteria for aspect-oriented programs. *Journal of Systems and Software*, 80(6):862–882, 2007.
- [44] S. Soares, P. Borba, and E. Laureano. Distribution and persistence as aspects. *Software - Practice & Experience*, 36(7):711–759, 2006.
- [45] A. van Deursen, M. Marin, and L. Moonen. A systematic aspect-oriented refactoring and testing strategy, and its application to JHotDraw. Report SEN-R0507, Stichting Centrum voor Wiskunde en Informatica, Netherlands, 2005.
- [46] R. Johnson et al. Spring - Java/J2EE application framework. Reference Manual Version 2.0.6, Interface21 Ltd., 2007.
- [47] M. Mortensen and R. T. Alexander. An approach for adequate testing of AspectJ programs. In *WTAOP'2005*, 2005.
- [48] W. E. Wong. On Mutation and Data Flow. PhD thesis, Dept. of Computer Science, Purdue University, USA, 1993.
- [49] Tao Xie and Jianjun Zhao, A Framework and Tool Supports for Generating Test Inputs of AspectJ Programs, *AOSD 06*, March 20–24, 2006, Bonn, Germany.
- [50] Romain Delamare, Benoit Baudry and Sudipto Ghosh, A Test-Driven Approach to Developing Pointcut Descriptors in AspectJ, *Proceeding, ICST '09 Proceedings of the 2009 International Conference on Software Testing Verification and Validation*.
- [51] Freddy Munoz, Benoit Baudry, Romain Delamare, Yves Le Traon, Inquiring the Usage of Aspect-Oriented Programming: An Empirical Study, European project DiVA (EU FP7 STREP), Software Maintenance, 2009. *ICSM 2009*.

# Modeling Data Transmission through a Channel Based on Huffman Coding and Encryption Methods

Eugène C. Ezin

*Institut de Mathématiques et de Sciences Physiques  
Unité de Recherche en Informatique et Sciences Appliquées  
Université d'Abomey-Calavi, République du Bénin  
eugene.ezin@imsp-uac.org*

**Abstract**—Data transmission through a secure channel requires the attention of many researchers. In this paper, on the basis of an alphabet of ciphers and letters, we propose a model for data transmission through a secure channel. This is achieved at two levels. First we associate each distinct symbol with a probability in the message to transmit. By doing so, we modify the well-known adaptive Huffman coding method. The obtained alphabet is used to construct the coded message to transmit through a cryptosystem. Therefore, the original message is coded and encrypted before its delivering. The proposed model is examined.

**Keywords**—component—Data compression, Huffman coding technique, encryption and decryption algorithms.

## I. INTRODUCTION

Data transmission occurs when there is a channel between two machines. To exchange a datum, an encoding must be chosen for the transmission signals. This basically depends on the physical medium used to transfer the data, the guaranteed data integrity and the transmission speed. Data transmission can be simple if there are only two machines communicating, or if only a single piece of data is sent. Otherwise, it is necessary to install several transmission lines or to share the line among several different communication actors. Transmitting a message sometimes requires a safe channel to stop an unauthorized person discovering its content.

So far, many codes have been used to represent letters and messages, among which Morse Code, ASCII code, and UNICODE are the most famous [1]. Codes with variable length including Huffman coding, are very useful in data compression field. It is worth noting that Huffman binary codes have the property that no character code is a prefix to any other character code. This is merely due to the fact that all the characters occur at only at the leaf nodes in the tree.

Applications of Huffman coding are pervasive in computer science. This coding scheme is not limited to encoding messages. Indeed, Huffman coding can be used to compress parts of both digital photographs and other files such as digital sound files (MP3) and ZIP files.

In the case of JPEG files, the main compression scheme uses a discrete cosine transform, but Huffman coding is used as a minor tool in the overall JPEG format. There are of course many other file compression techniques besides Huffman

coding, but next to run-length encoding, Huffman coding is one of the simplest forms of file compression. Huffman coding can be used effectively wherever there is a need for a compact code to represent a long series of a relatively small number of distinct bytes.

For instance in [2], Aljihit et al. present a compression/decompression scheme on selective Huffman coding for reducing the amount of test data. The difficulty observed to break the ciphered message shows that the proposed cryptosystem can be used as a secure channel for data transmission. That must be stored on a tester and transferred to each core in a system on a chip during manufacturing test. The request for electronic documents and services is increasing with the widespread use of digital data. Usually, electronic documents go through two separate processes: data compression to achieve low transmission cost, and ciphering to provide security. By using Huffman codes [3] we intend to keep information retrieval system functions like indexing and searching in the compressed file, which is not so easily possible with adaptative data compression algorithms. Huffman codes have some advantages, namely simplicity, speed, auto synchronization. Advantage means that it is possible to decode symbols in the middle of the coded text.

Another application of Huffman coding is the reduction of the peak to average in orthogonal frequency division multiplexing(OFDM) system [4] which is a frequency modulation technique for transmitting large amounts of digital data over a radio wave. OFDM works by splitting the radio signal into multiple smaller sub-signals that are then transmitted simultaneously at different frequencies to the receiver.

Klein et al. [5] analyzed the cryptographic aspects of Huffman codes used to encode a large natural language on CD-ROM and concluded that this problem is NP-complete for several variants of the encoding process [6]. Rivest et al. [7] cryptanalysed a Huffman encoded text assuming that the cryptanalyst does not know the codebook. According to them, cryptanalysis in this situation is surprisingly difficult and even impossible in some cases due to the ambiguity of the resulting encoded data. Data compression algorithms have been considered by cryptographers as a ciphering scheme [8].

Online banking is one of the most sensitive tasks performed

by general Internet users. Most traditional banks now offer online banking services, and strongly encourage customers to do online banking with *peace of mind*. Although banks strongly advertise an apparent *100% online security guarantee* typically the fine print makes this conditional on users fulfilling certain security requirements. In [9] Mohammad intended to spur a discussion on real-world system security and user responsibilities, in a scenario where everyday users are strongly encouraged to perform critical tasks over the Internet, despite the continuing absence of appropriate tools to do so.

In this paper we investigate the construction of a code associated with the 36 alphanumerical symbols. We define a random value probability for any symbol in the considered alphanumerical alphabet. This can be viewed as a dynamical adaptive Huffman coding procedure. The obtained symbols' codebook is used to code message to transmit through an encryption system. The paper is organized as follows. The section II presents some basic concepts. Section III describes techniques for codes construction. Section IV presents the Huffman algorithm. Section V describes the proposed model for data transmission. Section VI presents the conclusion and perspectives.

## II. CONCEPTS

In this section, we review some basic concepts that will be used in the paper.

### A. Alphabet

An alphabet  $\mathcal{A}$  is a finite and nonempty set of elements called symbols or letters. Throughout this paper, we will use the terms symbols or letters interchangeably.

We are familiar with some alphabets. For instance the Latin alphabet contains twenty six symbols from  $A$  to  $Z$ . The Roman digit alphabet is composed of  $I, V, X, L, C, D, M$ . The decimal alphabet contains the symbols  $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ . A binary alphabet contains the symbols  $0$  and  $1$ .

### B. Word

A word is a finite ordered combination of symbols belonging to a given alphabet. A vocabulary is usually constructed on an alphabet for the definition of words with a meaning.

A length of a word is the number of symbols it contains whatever they are different or identical.

A word without a symbol is called empty word and is denoted  $\epsilon$  with length equal to  $0$  by convention.

A set of all words on a given alphabet  $\mathcal{A}$  is denoted  $\mathcal{A}^*$ .

### C. Prefix of a word

Given two words  $w$  and  $u$  defined on an alphabet  $\mathcal{A}$ , the word  $u$  is a prefix of the word  $w$  or a left-factor of  $w$  if there exists a word  $v$  defined on the same alphabet such that  $w = uv$ .

### D. Concatenation operator

Given a couple of words  $u$  and  $v$ , the concatenated word of  $u$  and  $v$  is  $w$  defined by putting the second after the first. In other words,  $w = u \cdot v$  or for simplicity  $w = uv$ . The concatenation defines a monoid structure on  $\mathcal{A}^*$ . The words  $u$  and  $v$  are the factors of the word  $w$ . In consequence, a word can have many factorizations.

### E. Morphism

Given  $\mathcal{S}^*$  and  $\mathcal{A}^*$  two sets of words on alphabets  $\mathcal{S}$  and  $\mathcal{A}$  respectively, a mapping  $c$  from  $\mathcal{S}^*$  into  $\mathcal{A}^*$  is a morphism if it satisfies the following requirements:

- $c(\epsilon) = \epsilon$
- $\forall u, v \in \mathcal{S}^*, c(uv) = c(u)c(v)$ .

## III. CODING

Given a source alphabet  $\mathcal{S}$  and a target alphabet  $\mathcal{T}$ , a coding is an immersion morphism that satisfies the following

$$\forall u, v \in \mathcal{S}^*, u \neq v \implies c(u) \neq c(v).$$

A language is then a code if there is no word with two factorizations.

Given a set language  $L$ , one can use the Sardinas–Patterson algorithm to determine if  $L$  is a code or not. Before presenting the different steps of such an algorithm, let us first define the residual of a language and the quotient language.

### A. Residual language

Let  $L \subset \mathcal{A}^*$  be a language and  $u \in \mathcal{A}^*$  a word. We call a left-residual language of  $L$  by  $u$  the language denoted  $u^{-1}L$  defined by

$$u^{-1}L = \{v \in \mathcal{A}^* | \exists w \in L \text{ such that } w = uv\}$$

To put it briefly, the left-residual of a language  $L$  by  $u$  is composed of suffix of words of  $L$  for which the prefix  $u$  is deleted.

### B. Quotient language

Let  $L$  and  $M$  be two languages. A left-quotient language of  $L$  by  $M$  is the language  $M^{-1}L$  defined by

$$M^{-1}L = \bigcup_{u \in M} u^{-1}L.$$

In other words,  $M^{-1}L$  is the union of all left-residuals of  $L$  by words belonging to  $M$ . So  $M^{-1}L$  is the set of all suffixes of words in  $L$  that have a word belonging to  $M$  as prefix.

### C. Sardinas-Patterson algorithm

The Sardinas-Patterson algorithm is used to verify if a given language is a code [10]. Given a language  $L$  on an alphabet denoted  $\mathcal{A}$ , the following steps can be followed to check if such a language is a code.

One has to define the following sets  $L_n$ :

$$\begin{aligned} L_0 &= L \\ L_1 &= L^{-1}L - \{\epsilon\} \\ L_n &= L^{-1}L_{n-1} \cup L_{n-1}L \quad \forall n \geq 2 \end{aligned} \tag{1}$$



The process ends if one encounters an already  $L_i$  calculated or if  $\epsilon$  belongs to  $L_i$ .

If there exists  $L_i$  such that  $\epsilon \in L_i$ , then  $L$  is not a code. Otherwise  $L$  is a code.

As said above, the Sardinas-Patterson algorithm is a useful tool to check whether a given language is a code or not. A generalization of Sardinas and Patterson characterization to z-code can be found in [11]. However, a major shortcoming of this algorithm is that it does not say anything about the optimality of the code. The following section presents the Huffman algorithm for constructing an optimal code.

#### IV. HUFFMAN CODING

Huffman D. in [3] introduced its algorithm in 1951 when solving the problem of finding the most efficient binary code assigned by R. M. Fano. Indeed, Shannon [12] and Fano [13] have developed together coding procedures for the purpose of proving that the average number of binary digit required per message approaches the average amount of information per message. Their coding procedures are not optimal. The Huffman (or variable length coding) method is a lossless data compression algorithm based on the fact that some symbols have a higher frequency of occurrence than others. These symbols are encoded in fewer bits than the fixed length coding producing in average a higher compression. The idea behind Huffman coding is to use shorter bit patterns for more frequent symbols. To achieve that goal, Huffman algorithm needs the probability of occurrence of each symbol. These symbols are stored in nodes and then sorted in ascending order of their probability value.

The algorithm selects two nodes (children) with the smaller probabilities and constructs a new node (parent) with a probability value equal to the sum of the probabilities of its children. This process ends when the node with the probability equal to 1 is constructed. The result of this algorithm is a binary tree data structure with the root being the node with probability equal to 1, and the last nodes (leaves) represent the original symbols. The binary tree is then traversed to reach its leaves when a certain symbol is needed. Moving to a right child insert a 1 in the bit pattern of the symbol, while moving to left child insert a 0. The result is a unique bit pattern for each symbol that can uniquely decoded. This is due to the fact that no code word can be a prefix of any larger-length code word.

##### A. Optimal code

The concept of optimal code was introduced in order to reduce the space occupied by coded data. The definition of optimal code involves the use of the average length of a code.

Let  $\mathcal{S}$  be a source of  $n$  symbols with the frequency distribution  $f$  and  $\mathcal{C}$  a mapping code from  $\mathcal{S}$  into  $\mathcal{A}^*$ . The average length of the mapping code  $\mathcal{C}$  is defined by

$$\bar{n}_{f,\mathcal{C}} = \sum_{s \in \mathcal{S}} f(s) \times |C(s)|$$

in which  $|C(s)|$  represents the length of the word  $s$ .

A code  $\mathcal{C}$  of  $\mathcal{S}$  is optimal with respect to the distribution  $f$  if there does not exist another code  $\mathcal{K}$  such that  $\bar{n}_{f,\mathcal{K}} \leq \bar{n}_{f,\mathcal{C}}$ .

##### B. Presentation of the Huffman algorithm

The Huffman algorithm selects two nodes (children) with the smaller probabilities and constructs a new node (parent) with a probability value equal to the addition of the probabilities of its children. This process ends when the node with probability equal to 1 is constructed. The result of this algorithm is a binary tree data structure with the root being the node with probability equal to 1, and the last nodes (leaves) represent the original symbols.

Let  $\mathcal{S}$  be a source of  $n$  symbols with  $f$  the frequency distribution. We can assume that

$$f(s_1) \geq f(s_2) \geq \dots \geq f(s_n).$$

Let us consider another source  $\mathcal{S}' = (\mathcal{S} - \{s_{n-1}, s_n\}) \cup \{s_{n-1,n}\}$  with the following frequencies for its symbols

$$\begin{cases} f'(s) &= f(s) \text{ if } s \neq s_{n-1,n} \\ f'(s_{n-1,n}) &= f(s_{n-1}) + f(s_n) \end{cases}$$

Then the following mapping

$$f' : s \mapsto \begin{cases} f'(s) &= f(s) \text{ if } s \neq s_{n-1,n} \\ f'(s_{n-1,n}) &= f(s_{n-1}) + f(s_n) \end{cases}$$

defines an optimal coding set.

The variant we introduced in the Huffman algorithm is described in the subsection V-A. It consists of a dynamic and adaptive procedure to assign the probability value to each symbol in the alphabet.

#### V. DATA TRANSMISSION SYSTEM

We will describe in this section, the proposed coding method based on the Huffman coding algorithm. First we present the technique for obtaining the probability assign to each symbol.

##### A. Assignment of Probability Values to Symbols

Considering the alphanumeric alphabet  $\mathcal{A}$  as a source of the 36 symbols. The assigned random values to the symbols can be arranged into a  $6 \times 6$ -matrix. The following pseudo-code generates such random values.

```
#define ROWS 6
#define COLS 6
float val[ROWS][COLS], sum = 0.0f;

for(int i = 0; i < ROWS; i++){
    for(int j = 0; j < COLS; j++){
        val[i][j] = rand(0,1);
        sum = sum + val[i][j];
    }
}

for(int i = 0; i < ROWS; i++){
    for(int j=0; j < COLS; j++){
        val[i][j] = val[i][j]/sum;
    }
}
```

A sample of generated values is given in Table I.

We then assigned each probability value to each symbol in the alphabet. Without loss of generality, the result obtained

TABLE I  
A SAMPLE OF GENERATED VALUES FOR PROBABILITY ASSIGNMENT TO  
SYMBOLS.

0.0394	0.0135	0.0463	0.0383	0.0328	0.0341
0.0438	0.0264	0.0235	0.0464	0.0366	0.0017
0.0061	0.0463	0.0387	0.0317	0.0259	0.0134
0.0441	0.0466	0.0069	0.0020	0.0190	0.0022
0.0306	0.0076	0.0204	0.0410	0.0317	0.0047
0.0047	0.0469	0.0442	0.0451	0.0083	0.0398

is written in Table III to construct the adaptive and dynamic code based on Huffman algorithm. Like Huffman code, it has the prefix property that states for no character code is a prefix to any other character code. It is due to the fact that all the characters occur only at the leaf nodes in the tree. More information on the theory of codes can be found in [14].

TABLE II  
SOURCE SYMBOLS WITH THEIR CORRESPONDING FREQUENCIES.

<b>Symbols</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
Frequency	0.0469	0.0466	0.0464	0.0463
<b>Symbols</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
Frequency	0.0463	0.0451	0.0442	0.0441
<b>Symbols</b>	<b>8</b>	<b>9</b>	<b>A</b>	<b>B</b>
Frequency	0.0438	0.0410	0.0398	0.0394
<b>Symbols</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
Frequency	0.0387	0.0383	0.0366	0.0359
<b>Symbols</b>	<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>
Frequency	0.0341	0.0328	0.0317	0.0317
<b>Symbols</b>	<b>K</b>	<b>L</b>	<b>M</b>	<b>N</b>
Frequency	0.0306	0.0264	0.0235	0.0204
<b>Symbols</b>	<b>O</b>	<b>P</b>	<b>Q</b>	<b>R</b>
Frequency	0.0190	0.0135	0.0134	0.0083
<b>Symbols</b>	<b>S</b>	<b>T</b>	<b>U</b>	<b>V</b>
Frequency	0.0076	0.0069	0.0061	0.0047
<b>Symbols</b>	<b>W</b>	<b>X</b>	<b>Y</b>	<b>Z</b>
Frequency	0.0047	0.0022	0.0020	0.0017

### B. Presentation of the obtained optimal code

We construct an optimal code by assigning each of the 36 random values probability to each symbol in the order they appear. Without loss of generality, one can consider the probability values assigned in ascending order with respect to the symbols in the alphabet  $\mathcal{A}$ . The result code obtained when applying

### C. Proposed Data Transmission Channel

Once the original message is formed, the set of different symbols in such a message is considered with assignment of random value probability to each symbol we defined in subsection V-A. Huffman algorithm is then applied to compress data to be transmitted. Next, an encryption method is used to encrypt the message before delivering it to the receiver. The proposed model is shown in Fig. 1.

On the receiver's side, the user requires to know the key of the encryption method and the adaptive codebook constructed based on random value probability assignment. Figure 2 shows the process for decoding the received message.

TABLE III  
SOURCE SYMBOLS WITH THEIR CORRESPONDING CODEWORDS.

Symbols	Codewords	Symbols	Codewords
0	1010	I	01110
1	1011	J	01111
2	1100	K	10000
3	1101	L	10010
4	1110	M	000010
5	1111	N	000011
6	00000	O	010010
7	00010	P	100011
8	00011	Q	100110
9	00100	R	0100111
A	00101	S	1000100
B	00110	T	1000101
C	00111	U	1001110
D	01000	V	01001100
E	01010	W	01001101
F	01011	X	10011111
G	01100	Y	00111100
H	01101	Z	100111101

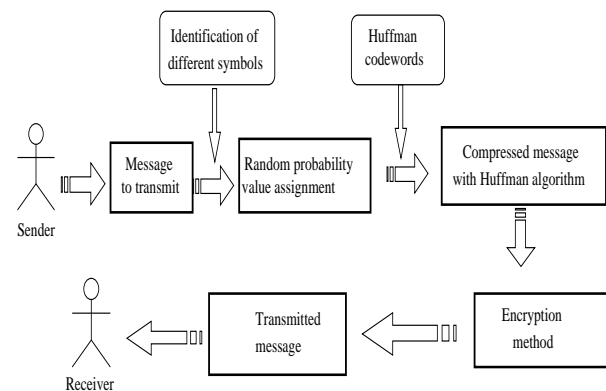


Fig. 1. Block diagram for transmitting the original message.

### D. Decoding Huffman Algorithms

While the Huffman encoding is a two pass problem, the decoding Huffman technique can be done in one pass by opening the encoded file and reading the frequency data out of it. It consists of:

- create the Huffman tree based on that information (The total number of encoded bytes is the frequency at the root of the Huffman tree.);
- read data out of the file and search the tree to find the correct char to decode (a 0 bit means go left, 1 go right for binary tree).

Many works have been done to construct decoding Huffman algorithms and their evaluated complexities. In [9] a parallel Huffman decoding algorithm for concurrent read and exclusive write, parallel random access memory which used  $N$  processors is proposed. Its complexity in time is  $O(\log n)$ . In [15] Pushpar et al. proposed a new approach to Huffman coding and in [16] a ternary for Huffman decoding technique. They used the concept of adaptive Adaptive Huffman coding based on ternary tree. This algorithm complexity is less than  $O(\log n)$ .

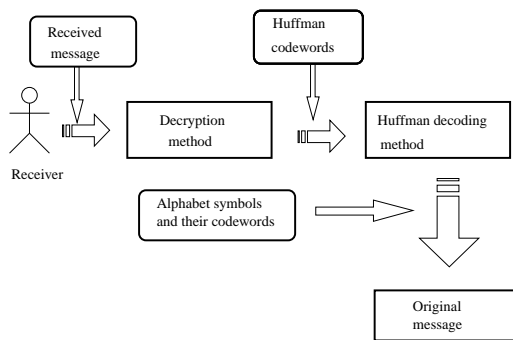


Fig. 2. Block diagram for decoding the received message.

## VI. CONCLUDING AND FURTHER WORKS

In this paper, we proposed a model for data transmission through a channel by introducing the use of dynamic and adaptive Huffman code with random values probability assignment. The resulting alphabet is used to code the message to transmit. This last is then encrypted based on encryption algorithm known in cryptography.

In our future works we will implement such a model by trying some algorithms in cryptography in order to evaluate its security level. This future work will include the evaluation of the confidentiality, integrity, and authentication.

## ACKNOWLEDGMENT

We thank anonymous reviewers for their review efforts.

## REFERENCES

- [1] S. Roman, *Coding and Information Theory*. Springer-Verlag, 1992.
- [2] J. Abhijit, "Computer-aided design of integrated circuits and systems," *IEEE Transaction*, vol. 22, no. 6, 2003.
- [3] D. A. Huffman, "A method of maximum and minimum redundancy codes," in *Proc. IRE*, 1952, pp. 1098–1101.
- [4] A. Ashraf, "Peak-to-average power ratio reduction in ofdm systems using Huffman coding," in *Proc. of World Academy of Science, Engineering and Technology*, vol. 33.
- [5] S. T. Klein *et al.*, "Storing text retrieval systems on cdrom: Compression and encryption considerations," in *ACM Transactions on Information Systems*, vol. 7, no. 3, 1989.
- [6] S. T. Klein and A. S. Fraenkel, "Algorithmica 12," 1989.
- [7] R. L. Rivest *et al.*, "On breaking a Huffman code," in *Proc. IEEE Transactions on Information Theory*, vol. 42, no. 3, 1996.

- [8] G. Simmons, *Contemporary Cryptology - the Science of Information Integrity*. IEEE Press, 1991.
- [9] M. Mohammad, "Authentication and securing personal information in an untrusted internet," Ph.D. dissertation, Carleton University of Ottawa, Ontario, 2009.
- [10] A. Sardinas and G. W. Patterson, "A necessary and sufficient condition for the unique decomposition of coded messages," *Convention Record of the Institute of Radio Engineers, National Convention. Part 8: Information Theory*, pp. 104–108, 1953.
- [11] M. Madonia *et al.*, "A generalization of sardinas and patterson algorithm to z-code," in *Theoretical Computer Science* 108, 1992, pp. 251–270.
- [12] C. E. Shannon, "A mathematical theory of communication," in *Bell System Technology Journal*, vol. 27, 1948, pp. 398–403.
- [13] R. M. Fano, "The transmission information," Research Laboratory of Electronics, MIT, Cambridge, Massachusetts, Tech. Rep., 1949.
- [14] J. Berstel and D. Perrin, *Theory of Codes*. Academic Press, 2002.
- [15] R. S. Pushpa and G. Madhu, "A new approach to Huffman coding," in *Journal of Computer Science*, vol. 4, no. 4, 2010.
- [16] R. S. and G. M. Pushpa, "Ternary tree and a new Huffman decoding technique," in *International Journal of Computer Science and Network Security*, vol. 10, no. 3, 2010.

## AUTHOR PROFILE



**Eugène C. Ezin** received his Ph.D degree with highest level of distinction in 2001 after research works carried out on neural and fuzzy systems for speech applications at the International Institute for Advanced Scientific Studies in Italy. Since 2007, he is a senior lecturer in computer science. He is a reviewer of Mexican International Conference on Artificial Intelligence. His research interests include high performance computing, neural network and fuzzy systems, signal processing, cryptography, modeling and simulation. He is currently in charge of the master program in computer science and applied sciences at the Institut de Mathématiques et de Sciences Physiques of the Abomey-Calavi University in Republic of Benin.

# ***PERFORMANCE COMPARISON OF MULTICAST ROUTING PROTOCOLS IN AD- HOC WIRELESS NETWORKS***

N.Veena<sup>1\*</sup>

Research Scholar,  
Dept. of CSE,  
Research Scholar,  
A N University,  
Guntur.  
veenanayani@gmail.com

A.VinayaBabu

Professor, Dept. of CSE,  
JNTU College of Engineering,  
Hyderabad,

K.Madhukar

Professor, Nizam College  
Osmania University  
Hyderabad

B.Sujatha

Lecturer in Computers  
Dept. of CSE  
Nizam College  
Hyderabad

B.Anandkumar

Research Scholar,  
Dept. of CSE,  
Rayalasima University,  
Kurnool.

S.Nagaprasad

Research Scholar,  
Dept. of CSE,  
Aacharya Nagarjuna University  
Guntur

**Abstract:** The majority of applications are in the areas where rapid deployment and dynamic reconfiguration are necessary and a wire line network is not available. These include military battlefields, emergency search and rescue sites, classrooms, and conventions, where participants share information dynamically using their mobile devices. Well established routing protocols do exist to offer efficient multicasting service in conventional wired networks. These protocols, having been designed for fixed networks, may fails to keep up with node movements and frequent topology changes in a MANET. Therefore, adapting existing wired multicast protocols as such to a MANET, which completely lacks infrastructure, appear less promising. Providing efficient multicasting over MANET faces many challenges, includes scalability, quality of service, reliable service, security, Address configuration, Applications for multicast over MANET. The existing multicast routing protocol do not addresses these issues effectively over Mobile Adhoc Networks (MANET).

## **1 Introduction**

In the last five years there is lot of research and of attention improved in the MANET network. To commutative in the MANET network we have two different types of protocols they are DSDV (Destination- Sequenced Distance vector) and DSR (Dynamic Sources Routing). For the last five years research there is no commercial implementation of DSDV, and also DSR is better performance then DSDV.

## **2 DSR (Dynamic Sources Routing)**

DSR simple algorithm based on source routing. As the described by (Mr. D. Johnson y. Hu & Maltz 2004), DSR is a simple and an efficient active routing protocol that designed especial for use of wireless mobile ad hoc networks. It is designed for high quality wireless network like more than 300 nodes. It incurs very low overhead but reacts very quickly to change the network. DSR protocol allows an ad hoc network to be self-organizing and self-

configuring without relaying on existing networks. Mobile nodes cooperative with each other for forwarding the packets over multiple hops between nodes that with in commutation range [1]. DSR protocol special designed for MANET applications, which are runs on wireless ad-hoc network. The very important feature of DSR is every data packet is follow the sources route stored in the header. The route gives clear address of each node through which packet should forwarded in order to reach the destination. Each node on the path has a routing role and must transmit the packet to the next hop identified in the source route. Each node maintains a route cache. When a node needs to send a data packet it first checks its route cache for a source route and destination, if route not found it attempts with discovery mechanism. A monitoring mechanism is called route maintains, it is used each operation along route [2]. Determining source routes requires accumulating the address of each device between the source and destination during route discovery. The accumulated path information is cached by nodes processing the route discovery packets. The learned paths are used to route packets. To accomplish source routing, the routed packets contain the address of each device the packet will traverse. This may result in high overhead for long paths or large addresses, like IPv6. To avoid using source routing, DSR optionally defines a flow id option that allows packets to be forwarded on a hop-by-hop basis. [3]

## **3 DSDV (Destination- Sequenced Distance vector).**

DSDV is one the early algorithm. This DSDV is one suitable for the small and less numbers of wireless ad hoc of nodes. Since there is no formal specification of this algorithm there is no commercial implantation. Many improved forms this algorithm have been suggested. DSDV

is regular update of the routing table and which uses a small amount of batter power and small amount of bandwidth even when the network is idle. When ever the topology of the network changes, a new sequence number is necessary before the network re-coverage. And DSDV is not suitable for high dynamic networks. As in all distance-vector protocols, this does not perturb traffic in regions of the network that are not concerned by the topology change. Destination-Sequenced Distance-Vector Routing (DSDV) is a table-driven routing scheme for ad hoc mobile networks based on the Bellman-Ford algorithm [4]. The main contribution of this algorithm is routing loop problem. Each entry in the routing table contains a sequence number, the sequence numbers are generally even if a link is present; else, an odd number is used. The number is generated by the destination, and the emitter needs to send out the next update with this number [5]. In DSDV a sequence number is linked to a distention node, and usually is originated by the node. The only case that non owner updates the sequence number of route is when it detects link brake on that route. Owner node is always uses an even sequence numbers and non owner always uses an odd numbers. With the addition of sequence numbers, routes for the same destination are selected based on the following rules: 1) a route with a newer sequence number is preferred; 2) in the case that two routes have a same sequence number, the one with a better cost metric is preferred. [4] DSDV uses both full and incremental updates of routing tables to reduce the routing overhead. The main advantage of proactive routing protocol is that a route to any destination is available even if it is not needed.

#### 4 Previous Work:

One of the popular wireless network architecture is mobile ad hoc network MANET which can be deployed almost any environment, without any underlying backbone and infrastructure. Mr. Rafi U Zamam is proposed an efficient DSDV protocol for Ad hoc network. Eff-DSDV overcomes the problem of stale routes, and thereby improves the performance of regular DSDV. This proposal protocol is implemented in the NCTUcs simulator for the comparisons of DSDV and DSR protocol. He considered the performance issues of Packet-delivery ratio, and the end to end delay of packets, dropped packets, and routing overhead. [5]

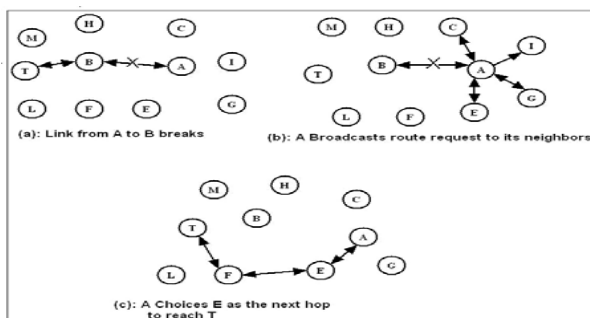


Figure: Example of linking nodes

The result of the simulation indicates that performance is better of standard DSDV. In this analysis they found that if there is it is better efficiency when the nodes range from 10 to 30. If the range is increased to beyond 35 the performance of the proactive protocol degenerates due to that the fact of a lot of control packets are generated in the form of incremental or full dump of the routing update/table in the sub net mask.

Even they observed that the proposed protocol is even better performance comparing regular DSDV protocol and verifying better speed in mobile ad hoc network. Even they observed that the better performance of proposed protocol comparison then DSR protocol like packet delivery of ratio and end to end delay. The region of the performance to set drop at 25 nodes at and in some case for nodes speed is due to verging sources and destination nodes and the placement of barrier in the network topology. In the feature, the performance comparison can be done with proposed protocol and the other class of the ad hoc routing protocol with the different simulation parameter and matrices. After their analysis they got clear information that the proposed protocol is even better performance and better end to end delay comparison of regular DSDV protocol.

Table 1: Route update table at host

Neighbor	No. of Hops	Via node	Update time
C	2	H	1765
E	2	F	1860
G	3	E	1050
I	3	A	805

The proposed protocol is designed by Mr. Bikas. The objective of Mr. Bikas is implementing two different DSDV and DSR routing protocols by using network simulators and run those protocols in different nodes. The compared two routing protocols are studies in different senior on the basis of packet delivery ratio and routing load. [6] DSDV is proactive gateway discovery algorithm where the gateway periodically broadcasts a gateway advertisement message which is transmitted after expiration of the gateways timer. DSR is reactive gateway discovery algorithm where a mobile device of MANET connects by the gateway only when it is needed. He implemented the Destination Sequence Distance Vector Dynamic Source routing protocols in TCL and integrated modules with network simulator 2. The performance comparison of these protocols is measured with packet end to end delivery and packets end to end delay. Simulation were carried out with identical topology and running different protocols on the mobile node. The performance result indicates that the DSR is better then DSDV protocol. It is also they observed that the performance is better when PAUSE time is low. And the Dynamic Source Routing delivery is better even when the network is higher and each other. And they observed that DSR is a Reactive gateway discovery algorithm where a mobile device of MANET connects by gateway only when it is needed. There is performance depends up on different

senior. The protocol is proposed by Mr. Jean-pierre.guedon, z.issaka and Salima Hamma. The proposed is performance compression between proactive and reactive routing protocols by using Network simulator 2. They have considered that qualities and quantities criteria. They considered the first one is operation, loop-freedom, security, sleep period operation and the second is second are used to assess performance of different routing protocols presented in this paper. We can list end-to-end data delay, jitter, packet delivery ratio, routing load, activity distribution. [7] The performance analysis is for proactive and reactive routing protocol by using network simulators 2. After this analysis they observed that the proactive routing protocol is better performance for CBR sources given that it guarantees lowest delay and jitter and it consumes more bandwidth. Periodically the proactive protocol sends routing packets to discover and to maintain routes to all destinations. The number of delivered packets decreases when the numbers of nodes are increases. For the less number connections the packet delivery ratio is 53%. The reactive routing protocols are more adapted for data services. They guarantee the 80% packet delivery for the more than 60 connections. There is no clear winner among the proactive and reactive routing load delay and jitter and quite identical. The proposed protocol is implemented by Mr. Samyak Shah<sup>1</sup>, Amit Khandre<sup>2</sup>, Mahesh Shirole<sup>3</sup> and Girish Bhole. The project implanted for the comparison of on demand reactive routing protocol of DSR along with the AODV with tradeseanal proactive DSDV routing protocol. A simulation model with MAC and physical layer models is used to study interlayer interactions and their performance implications. On-demand routing protocols, AODV and DSR performance is better than table driven DSDV routing protocol. [5] The AODV and DSR having similar behavior, but differences in the protocol mechanics can lead to significant performance differentials. And a verity of workload and scenario, as characterized by the mobility, load and size of the ad hoc network were simulated. The performance differential analyzes varying network load mad mobility and network size. [8] The project performance is compared for the DSDV and DSR and AODV protocols by using network simulator 2. They observe that simulation is that for application oriented matrices such as a packet delivery fraction and delay AODV, outperforms DSR in more "stressful" situation. They found that when the smaller number of nodes and lower load or mobility, with widening performance gaps with increasing stress. More load, higher mobility. Any ho they found that DSR produces less routing load comparing with AODV protocol. The poor performance of DSR mainly attributed to aggressive use of caching and lack of any mechanism to expire stale routes or determine the freshness of routes when multiple choices are available. And they found that the aggressive cache helps DSR to make low routing loads and make better performance. They found that better performance of DSR and AODV comparing with DSDV. And they found that when the network more than 30 nodes the DSDV routing protocol is more routing load on the network, and it is

observe that the DSR having less routing load on network. And they observe that the comparing with DSR better performance then DSDV routing protocol. DSR and AODV both use on-demand route discovery, but with different routing mechanics. In particular, DSR uses source routing and route caches, and does not depend on any periodic or timer-based activities. DSR exploits caching aggressively and maintains multiple routes per destination. AODV, on the other hand, uses routing tables, one route per destination, and destination sequence numbers, a mechanism to prevent loops and to determine freshness of routes. [8]

#### CONCLUSION:

In this paper I compared of DSDV with on-demand routing protocol DSR. With these reviews I found that DSR is better performance then DSDV. And second is that when for less no of wireless nodes in the network the performance of DSDV is better, while wireless nodes increases in the network the performance goes poor for DSDV routing protocol. The performance of DSDV is poor in the MANET because of routing load on network. DSR is better performance comparing to DSDV protocol, because of DSR having less routing load on network. Another thing that they observed that when no of wireless nodes are increased there is poor performance for both protocols. But DSR is better performance then DSDV.

#### REFERENCES:

- [1] Johnson, D., Y. Hu & Maltz, D. 2004, The Dynamic Source Routing Protocol (DSR) for Mobile Ad hoc Networks for IPv4, RFC 4728.
- [2] Feeney, L.M. 1999, A taxonomy for routing protocols in mobile ad hoc networks, Swedish Institute of Computer Science.
- [3] D. Johnson, D. Maltz, and J. Broch. DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks. In *Ad Hoc Networking*, C. Perkins (ed), Chapter 5, pages 139–172, 2001.
- [4] C.E. Perkins & P. Bhagwat, "Highly Dynamic Destination Sequence-Vector Routing (DSDV) for Mobile Computers", *Computer Communication Review*, vol. 24, no.4, 1994, pp. 234-244.
- [5] Mr.Rafi U Zamam "An Efficient DSDV Routing Protocol for Wireless Mobile Ad Hoc Networks and its Performance Comparison  
<http://www.computer.org/portal/web/csdl/doi/10.1109/EMS.2008.11>"
- [6] Mr Bikas: Implementing and comparing dsr and dsdv routing protocols for mobile ad hoc networking."  
<http://ethesis.nitrkl.ac.in/1131/1/bikas.pdf>  
<http://www.eurojournals.com/ejsr.htm>
- [7] Jean-pierre.guedon , z.issaka and Salima Hamma hal-00275330, version 1 - 23 Apr 200. "<http://www.eurojournals.com/ejsr.htm>"
- [8] Samyak Shah<sup>1</sup>, Amit Khandre<sup>2</sup>, Mahesh Shirole<sup>3</sup> and Girish Bhole  
<http://tifac.velammal.org/CoMPC/articles/35.pdf> page 167 to 171.



## A Proposed Ontology Based Architecture to enrich the data semantics syndicated by RSS techniques in Egyptian Tax Authority

Ibrahim M El-Henawy<sup>1</sup>  
<sup>1</sup>Faculty of computer and Informatics  
Zagazig University  
Zagazig, Egypt  
Henawy2000@yahoo.com

Mahmoud M Abd El-latif<sup>2</sup>  
<sup>2</sup>Faulty Of Computer and Information  
Mansoura University  
Mansoura, Egypt  
drmmlatif@yahoo.com

Tamer E Amer<sup>2</sup>  
Tameramer1@yahoo.com

**Abstract**—RSS (RDF site summary) is a web content format used to provide extensible metadata description and syndication for large sharing, distribution and reuse across various applications; the metadata provided by the RSS could be a bit to describe the web resource; this paper provides a framework for making the RSS not only just for syndicating a little information about news but also for further classification, filtering operations and answering many questions about that news by modeling RSS ontology. The proposed architecture will be applied to handle announcements in the Egyptian Tax authority.

**Keywords**- Semantic Web - RDF - Ontology - RSS - OWL - Protégé - Egyptian Tax

### I. INTRODUCTION

Egyptian tax authority consists of what is called 39 Tax Regions distributed all over Egypt that manages 227 tax offices [22]. All tax offices are connected via a huge computer Network on a single domain Called GTAX Domain managed by the central management of computer in Cairo. There are 14 IT (Information technology) branches to support the IT works in all tax regions. Besides the huge computer network; the Tax authority uses a huge IP telephone network that uses the VoIP (Voice over IP) technology to support communications between remote offices.

The idea of centralization makes a great challenge here; For example when the central management of computer in Cairo wants to announce for a specific event, meeting or a new version of specific application in the authority, it put a written announcement (.doc format) in the main FTP server and telephone all the 14 IT branches using IP telephone and then the 14 IT branches call the rest of 227 remote tax offices. It is a very time consuming manual announcing protocol; but using the RSS technique to syndicate data published by different places will facilitate the data exchange between them.

RSS can be found as acronym for RDF Site Summary; it is an RDF (Resource description Framework) vocabulary that provides a lightweight multipurpose extensible metadata to describe and syndicate any information consists of discrete items [1, 15 and 16]; hence It allows the key elements of websites, such as headlines, to be transmitted, when devoid of all elaborate graphics and layouts, such minimalist headlines are quite easily incorporated into other websites.

Besides the ability of RSS to solve many problems that web masters face such as increasing traffic, and gathering and distributing news, RSS can also be the basis for additional content distribution services. Regardless of the speed of

looking at many different sites in a single coherent hole, the democratic manner in news distribution that enables the user to choose the feed he wants; making him the potential news provider, can be considered the most efficient benefit in using RSS [2].

Many advantages can be achieved by using RSS, but what is noticed that all the data gathered in the RSS file is shown directly by any RSS aggregator. What about if someone wants to classify the data presented in RSS? For example; if the training management of Tax authority announces for a training course in "Soft Skills"; does this announcement belongs to specific department or for all?, does it for specific tax region according to specific schedule or for all? What if someone wants to know some information about the writer or publisher of the published article? It's obvious that there are many questions in the chain and the little Metadata description presented in the RSS technology did not have the ability to give answers for the questions chain.

Semantic web extends the current web by giving information published on the web a well-defined meaning, better enabling computer and people to work in corporation [3]. To make the RSS has the ability to answer the questions above; the word "well-defined meaning" should exist in the perspective of RSS; it is noticed that it may not be expressed via terminologies in RSS. The only way to express "well-defined meaning" in RSS is to extend the RSS itself by enabling it to link and interact with other ontologies; thus enriching the semantics that are provided by RSS.

The contribution of this paper is dealing with data published by RSS as domain ontology and enables it to interact with other vocabularies such as Dublin core Metadata, FOAF (Friend Of A Friend) ontology and tax ontology. This way enables us to make further operations about RSS data such as classifications, reasoning or answering the above questions chain. The presented ontology is modeled by Protégé.

The outline of this paper is as follows: providing a background of Egyptian tax authority and the current way of announcement in section 1. Section 2 illustrates what is the RSS and how it is related to RDF. The proposed architecture and the implementation of the RSS ontology are presented in section 3; finally we conclude this paper in section 4.

### II. LITERATURE ON RSS

RSS file is XML based syntax; it has xml/application MIME (Multi-purpose Internet Mail Exchange) type. The

extension of the file of RSS version 1.0 is preferable to be (.rdf). Care should be taken here because RSS will be discussed in the scope of RDF; RSS 0.9 and 1.0 is the only specification standard of RSS that uses RDF vocabularies the other specification (RSS 0.91, 0.92, 0.93, 0.94, and 2.0) does not [1, 16]; they are more basic XML implementation. Its file uses mainly the following two namespaces as two attributes within <rdf:RDF> tag:

- xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
- xmlns=http://purl.org/rss/1.0/

#### A. The anatomy Of RSS file

Each RSS file consists of a single channel that contains the information gathered from many different sites; it is represented by the <channel> element. The attribute rdf:about is used with in <channel> tag to describe the location and the name of the RSS file.

Some required tags within the channel element can be used to describe the channel itself such as

- <title> element to describe the title of the channel.
- <link> element to describe the URL of the parent site or the news page.
- <description> element to provide a brief description of the channel contents, function, etc.

As shown in figure.1; the channel contains number of items (<items>) listed in an ordered collection described by the RDF container <rdf:seq>. The items listed in the channel will be described outside the channel after the closing </channel> tag using the above <title>, <link> and <description> tags. The following block diagram illustrates the anatomy of the RSS file.

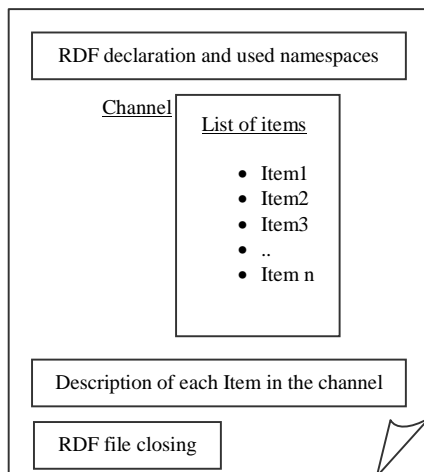


Figure.1 Anatomy of RSS File

#### B. The relationship of RSS to RDF

Earlier versions of RSS did not include any RDF vocabularies; it is just a syntactic XML representation of the published news. Although XML is a universal Meta language for defining Markup [7]; it is worth mentioned that XML has some deficits, for example it does not provide a satisfactory

semantics representation that is embedded in statements; there is no standard way exists to assign meaning to the nesting of the XML elements.

Expressing the RSS 1.0 in a language described in RDF concepts and abstract syntax makes it conforms to RDF/XML syntax specification that has a precise formal semantic defined in RDF semantics; thus easy interoperability with other RDF Languages and obviously can be read and processed by machines [13].

The foundation of RSS 1.0 serves the purpose of this paper that intends to extend the RSS 1.0 to be used outside of strict news and announcements syndication by focusing on a generic means of structured metadata exchanging [4] and how it can incorporate with other RDF ontologies by providing a simple modular extension mechanism to accommodate new vocabularies.

### III. SYSTEM ARCHITECTURE

The main purpose of the proposed architecture is to extend the RSS data gathered from different resources to exceed just syndication purpose by making an ontology that can interact with other ontologies to have a tight and well-defined metadata about the news and announcements. It will make a collaborative space that makes everything is linked. Classification operation can be done as well as many questions can be answered.

The framework presented in this research can be considered as integrated semantic web architecture. The word “integrated” refers to that this architecture consists of more than one component, each one has a specific task, and the word “semantic” means that, this architecture is based on semantic web technologies to make the presented ontology. Figure.2 shows the schematic diagram for this architecture

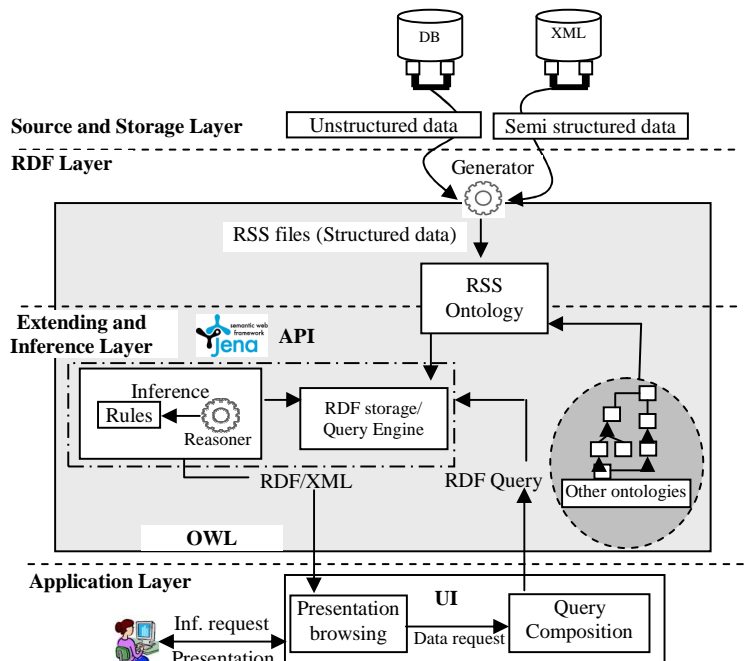


Figure.2 The schematic diagram of the architecture

Identify applicable sponsor/s here. (sponsors)

#### A. The Ontology of RSS data

Ontology is the heart of semantic web applications. The definition of the ontology [5] is the explicit and formal specification of conceptualization of a domain of interest. It is increasingly seen as a key technology for enabling semantics-driven knowledge processing. Communities establish ontologies, or share conceptual models, to provide a framework for sharing a precise meaning of symbols exchanged during communication and enable the programs to reason about different worlds and environments; they enable us to say “our world looks like this” [6, 10].

The presented ontology will be modeled by Protégé [17]; it is one of the most famous and widely used ontology editing environments [22, 23]. The conceptual Model of the proposed ontology will consist of the hierarchy of all Classes, subclasses, properties, sub properties and how it related to each other. The ontology is written in OWL (Ontology Web Language); it is a very powerful tool for describing complex relationships and characteristics of the resources. OWL allows rules to be asserted to classes and properties. Rules represent the logic that enhances the ontology language [7]; this will help when it is applied to a set of facts to infer new facts that are not explicitly stated.

Looking to the RSS data as ontology will help during searching for a concept to easy locate, not only the concepts but also the other concepts that are semantically related to it. Although RSS 1.0 is basically expressed in the language that is described in RDF concepts and abstract syntax; and RDF provides an ideal encoding to make available ontologies to semantic web applications; it offers a limited set of semantic primitives and cannot therefore meet the requirements of a markup language for the semantic web. So extending the RSS semantics by adding more primitives encoded in OWL to offer appealing inference capabilities will form a very tight defined vocabularies that describe the concepts in the ontology, and also exert significant influence on searching information about the concepts; the degree to which terminologies are semantically precise has a direct impact on the degree to which relevant information can be found [8, 9].

RDF schema should be considered when talking about the RSS 1.0 ontology because it shapes and describes the ontology of RSS 1.0. It is described in formal language in RDF schema of RSS 1.0 [18]. It consists of the following classes and attributes summarized in table I and table II

TABLE I. THE CLASS SPECIFICATION OF RSS IN RDF SCHEMA

Class	Definition	URL
Channel	An RSS information channel	<a href="http://purl.org/rss/1.0/channel">http://purl.org/rss/1.0/channel</a>
Image	An RSS Image	<a href="http://purl.org/rss/1.0/image">http://purl.org/rss/1.0/image</a>
Item	An RSS Item	<a href="http://purl.org/rss/1.0/item">http://purl.org/rss/1.0/item</a>
TextInput	An RSS text Input	<a href="http://purl.org/rss/1.0/textinput">http://purl.org/rss/1.0/textinput</a>

TABLE II. THE PROPERTY SPECIFICATION OF RSS IN RDF SCHEMA

Property	Definition	URL	Sub Property Of
Items	list of rss:item elements that are members of the subject channel	<a href="http://purl.org/rss/1.0/items">http://purl.org/rss/1.0/items</a>	
Title	A descriptive title for the channel	<a href="http://purl.org/rss/1.0/title">http://purl.org/rss/1.0/title</a>	Dublin core title element ; dc:title [19]
Link	The URL to which an HTML rendering of the subject will link	<a href="http://purl.org/rss/1.0/link">http://purl.org/rss/1.0/link</a>	Dublin core identifier element; dc:identifier [19]
url	The URL of the image to be used in the 'src' attribute of the channel's image tag when rendered as HTML.	<a href="http://purl.org/rss/1.0/url">http://purl.org/rss/1.0/url</a>	dc:identifier
Description	A short text description of the subject	<a href="http://purl.org/rss/1.0/description">http://purl.org/rss/1.0/description</a>	Dublin Core description element dc:description [19]
Name	The text input field's (variable) name	<a href="http://purl.org/rss/1.0/name">http://purl.org/rss/1.0/name</a>	

Figure.3, 4 represents the Node and Arc diagram for the classes and properties in the table I and table 2.

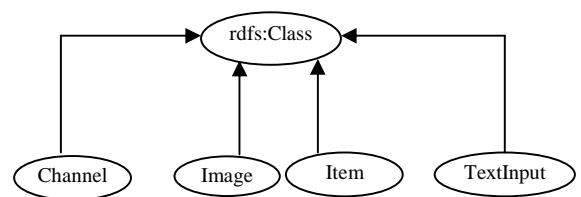


Figure.3 Node and Arc Diagram for RSS Classes

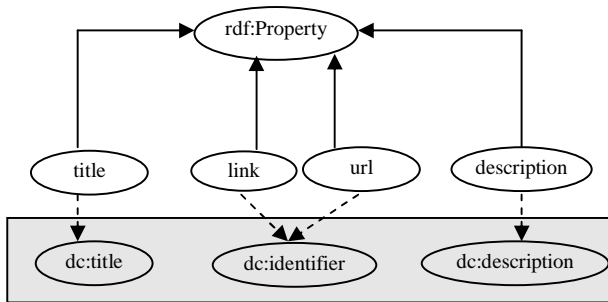


Figure.4 Node and Arc Diagram for RSS Properties

Solid lines in figure.3 and figure.4 represent the `rdf:type` relationship. The dashed lines in figure.4 represent the `rdfs:subPropertyOf` relationship.

### B. RSS ontology development and Implementation

RSS ontology is used to capture knowledge about RSS domain by describing the concepts and relationships that are held between those concepts. There are many different ontology languages that provide different facilities. RSS ontology will be modeled by OWL [20]; the most recent standard of ontology languages from W3C (World Wide Web Consortium). It has richer set of operators such as And, Or and Not. It is easy with OWL to define and describe concepts. It is easy to build complex concepts up in the definition of simpler concepts.

As logical model of OWL allows using reasoners; RACER reasoner [21] will be used to check whether or not all of the statements and definitions in the ontology are consistent. It also used to recognize which concepts fit under which definitions thus maintaining the correctness of the hierarchy.

#### 1) RSS Ontology Design

Building ontologies is divided into three steps; ontology capture, ontology coding and integrating with other ontologies [23]. RSS ontology consists of three main components; Classes to represent the concepts within the ontology, properties to relate the classes and individuals that belong to those classes. The novelty of this RSS ontology is that all concepts are tied using annotating properties and rules of Description Logic thus providing a rich concept definition.

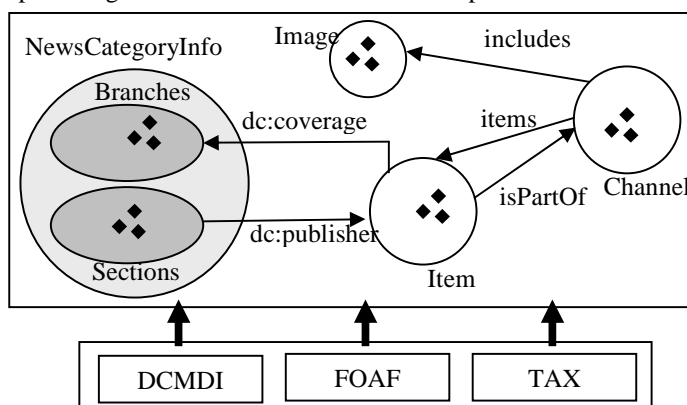


Figure.5 RSS Ontology Classes Relationship- other properties that describe item are omitted for better readability

Figure.5 illustrates how the components are related to each other. It is considered when developing this RSS ontology to follow ontology engineering methodology [11, 23] which consists of the following steps

- 1- Determine Domain and Scope
- 2- Consider reusing existing ontology
- 3- Enumerate terms in the ontology
- 4- Define Classes and hierarchy
- 5- Define Properties
- 6- Refine Properties
- 7- Create Instances
- 8- Consistency checking using reasoners

The domain and scope of RSS ontology was pretty clear from the beginning. The goal is to design RSS ontology that enables rich Metadata in order to be easy to query, make auto classifications and getting new concepts for RSS data itself according to the submitted rules.

Regarding to reusing of ontologies; RSS ontology reuses vocabularies of Dublin Core Metadata (DCMD) [19]. The goal of using Dublin core vocabularies in this ontology is to provide further description for resources. The resources can be further described by any other vocabularies made by the author, but the concept of using standardized descriptive metadata that the Dublin core addresses will be a powerful mechanism to improve information retrieval for specific applications [12]. Further description of RSS ontology concepts will make the aim of this ontology not only for syndicating news headlines and associated metadata, but also for transmitting complete structure datasets [4].

As every management in Tax authority has the right to publish its own news and announcements; each set of syndicated data will be packed in a single entry to be an item in the **Channel** class according to from where it is published, for example we can find `mansoura_branch_channel`, `Central_management_channel`, etc. **Channel** class is related to **Item** Class via (`items`) property that refers to a list of items belongs to the specific channel. The object property **includedIn** will be added to refer to the channel that specific item belongs to; it is inverse property of the `items` object property. In the specification of RSS 1.0 there is no relation between **Image** Class and **Channel** Class so the object property **includes** is added; it is a subproperty of the `dc:relation` Dublin core property to refer to the relation between the channel and the image classes. Making it a subclass of the `dc:relation` property makes sense: Consider an application which does not know our ontology; even if it does not know the meaning of **includes**, it still can infer that something has a relation to other one via this “includes” relationship. The title, description and link data type properties that describe the Item and Channel classes are defined as listed in table2.

The **NewsCategoryInf** class (News Category Information) is added to the core specification which represents the concepts of classification of RSS data according to the publisher it comes from or the destination it is delivered to. As in figure.5 it is a super class to other two subclasses **Sections** and **Branches**, It relates to the items that

hold the news by two main Dublin Core object Properties; dc:publisher and dc:coverage. The property dc:publisher refers to the publisher that items come from and dc:coverage refers to the destination to which the item is delivered. Figure.6 will illustrate ontology hierarchy generated by protégé.

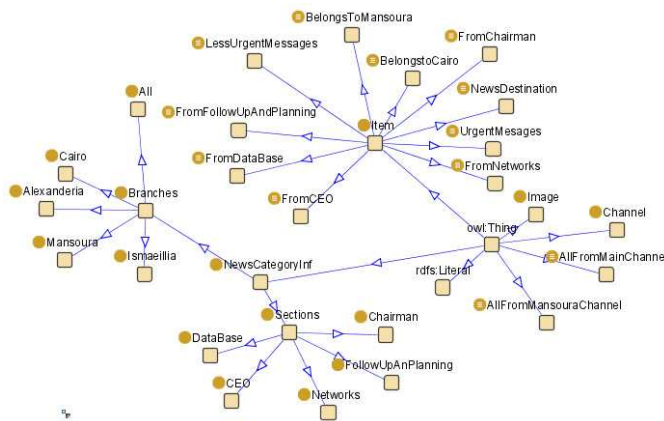


Figure.6 Ontology hierarchy – generated by protégé

## 2) Inference rules and reasoning the Ontology

The power of using OWL in the news syndication is that it can provide the ability to add a set of rules that can be asserted to the classes to infer new concepts that are not explicitly stated; thus enhancing the ontology. Regarding semantic web stack [7] we can find that the logic layer based up on ontology layer, it could not be used up on the layer of RDF and RDF schema layer; so rules can not be used with the existing standard of RSS version 1.0.

In RSS ontology we can infer list of concepts that are not explicitly stated from the existing knowledge. Table III lists a set of inferred classes such as the messages that belong to particular branch (mansoura, Cairo), messages that were published from specific section or branch (mansoura, Cairo, database, Networks, chairman or CEO (Chief executive Officer)), the destinations that the news cover. More Knowledge can be inferred such as what are the urgent and less urgent messages.

TABLE III. INFERRED CLASSES AND THEIR RULES

Inferred Class	Meaning	Rule
AllFromMainChannel	All news from the channel of the central management in Cairo	isPartOf has Mansoura_branch_channel
AllFromMansChannel	All news from mansoura channel	isPartOf has the central management channel
BelongsToCairo	Messages belongs to Cairo branch	dc:coverage some (Cairo or All)

Inferred Class	Meaning	Rule
BelngsToMansoura	Messages Belongs to mansoura branch	dc:coverage some (mansoura or All)
FromCEO	Messages from CEO	dc:publisher some CEO
FromChairman	Messages from Chairman	dc:publisher some Chairman
FromDatabase	Messages from DB section	dc:publisher some Database
FromNetworks	Messages from Network	dc:publisher some Networks
FromFollowUpAndPlanning	Messages from follow up and planning	dc:publisher some FollowUpAndPlanning
UrgentMessages	Urgent messages	dc:publisher some (CEO or Chairman)
LessUrgentMessages	Less urgent messages	not UrgentMessages
NewsDestinations	Destination of the messages	dc:coverage some Branches

The rules are asserted as necessary and sufficient conditions [14]. To make those classes “defined classes”. It basically means that any item that has at least one relationship to specific class along the property in the rule can be inferred as a member in the inferred class and in the same time any individual in the inferred class should meet that rules asserted to the class.

This rules do not only used to define the members belong to specific class but also used by the RACER reasoner to indicate the new inferred hierarchy. New semantics are added to classes, such as which of those classes are sub class of the other. It used to redefine the hierarchy. Figure.7 shows the asserted classes versus the inferred classes reasoned by the RACER reasoner in the protégé.

The asserted hierarchy appears in the protégé left panel. All defined classes are subclasses of the class **Item** except the two classes **AllFromMainChannel** and **AllFromMansChannel**. After using RACER reasoner to classify taxonomy; we find the inferred hierarchy in the right panel in a different form. It is clear that all defined classes become subclasses from Item Class and categorized to three main categories; **UrgentMessages** category that includes FromCEO and FromChairman classes, **LessUrgentMessages** (the complement of UrgentMessages) that includes FromDatabase, FromFollowUpAndPlanning and FromNetworks Classes and the final category is **NewsDestination** that includes BelongstoCairo and BelngsToMansoura classes. It is clear that the two classes AllFromMainchannel and AllFromMansChannel are not categorized in any of the three categories because it is not logically to be subclass of any of them. So they just categorized as a subclass of Item Class.



Future studies can be done towards designing and constructing of knowledge extracting system based on the proposed ontology.

## REFERENCES

- [1] Frank Manola, Eric Miller, "RDF Primer", <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- [2] Kevin Curran, Sheila Mc Kinney, "Delivering Audio and Video with Rich Site summary Enclosures" JOURNAL OF MULTIMEDIA, VOL. 1, NO. 1, APRIL 2006
- [3] Tim Berners-Lee, "The semantic web made easy", Scientific American, May 2001. <http://www.w3.org/RDF/Metalog/docs/sw-easy>.
- [4] Tony Hammond, Timo Hannay and Ben Lund, "The Role of RSS in Science Publishing: Syndication and Annotation on the Web", Nature Publishing Group.
- [5] John Davies, BT UK - Rudi Studer University of Karlsruhe, Germany - Paul Warren, BT, UK Semantic Web Technologies - Trends and Research in Ontology-based Systems. (2006) - DDU
- [6] Edd Dumbil, "building the semantic web", 2001, <http://www.xml.com/pub/a/2001/03/07/buildingsw.html>
- [7] Grigoris Antoniou and Frank van Harmelen, "A semantic web primer", MIT, 2004
- [8] Felipe M. Villoria, Oscar Díaz and Sergio F. Anzuola, "Powering RSS Aggregators with Ontologies: a case for the RSSOwl aggregator", The Onekin Group, University of the Basque Country, Spain
- [9] Antonio Sanfilippo, Battelle/Pacific Northwest Division, USA, Ansgar Bernardi, Deutsches Forschungszentrum für Künstliche Intelligenz, DE, Ludger van Elst, Deutsches Forschungszentrum für Künstliche Intelligenz, DE, Luis Sánchez Fernández, Universidad Carlos III de Madrid, ES, Michael Sintek, Deutsches Forschungszentrum für Künstliche Intelligenz, DE, "Position Paper: Integrating Ontologies for Semantic Web Applications", ENABLER/ELSNET Workshop on International Roadmap for Language Resources, 2003
- [10] Simone A. Ludwig and S.M.S. Reyhani, "Semantic approach to service discovery in a Grid environment", journal of web semantics, 2005.
- [11] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating our first ontology." <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>.
- [12] Diane Hillmann, "Using Dublin Core", [dublincore.org](http://dublincore.org), 2005
- [13] Danny Ayers, "Extending RSS", 2003, <http://www.xml.com/lpt/a/1254>
- [14] M. Horridge, H. Knublauch, A. Rector, R. Stevens, and C. Wroe, "A practical guide to building owl ontologies using the protg-owl plugin and co-ode tools.", <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf> (viewed 28/06/06), 2004.

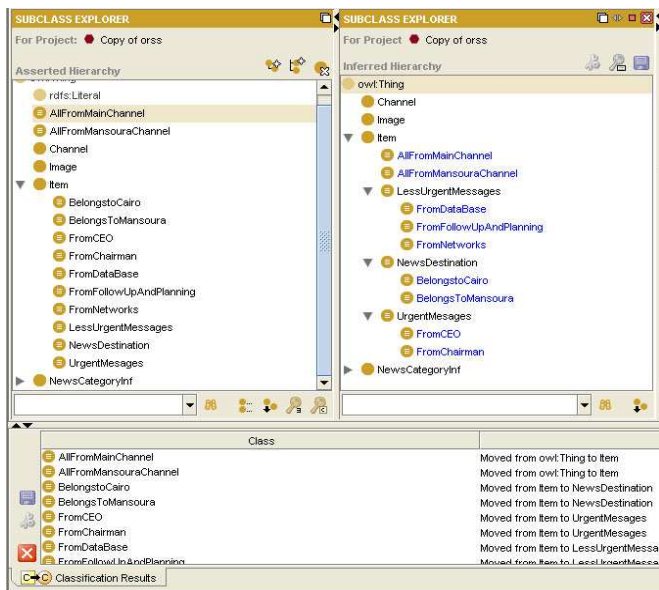


Figure.7 Asserted versus Inferred Class Hierarchy

The third panel in the bottom shows what happened to the defined classes to form the inferred hierarchy. This panel provides a facility to assert certain change by taking the superclass-subclass relationship that has been found by the reasoner and manually put them into the asserted hierarchy [14] but it is a bad idea to do whilst the ontology is being developed.

When we consider Classes **UrgentMessages** and **LessUrgentMessages** we find logically that object or individual can not be an instance of both classes in the same time so those classes should be disjoint classes. While **UrgentMessages** Class includes only messages from chairman or CEO (**FromChairman** and **FromCEO** classes); the **LessUrgentMessages** will not be a super class of any subclasses except they are disjoint with the classes belongs to **UrgentMessages**; this is because reasoning in OWL (description logic is based on open world assumption); that means in RSS ontology, we can not assume that individual is not a member of a particular class because it has not been asserted to be a member in that class. So the **FromDatabase**, **FromNetworks** and **FromFollowUpAndPlanning** should be disjoint with the two classes belong to **UrgentMessages** Class; this explicitly said that those classes couldn't be subclasses of **UrgentMessage** Class.

## IV. CONCLUSION AND FUTURE WORK

In this paper we introduced an ontology based on OWL development to describe the concepts about RSS data syndicated by different managements in Egyptian tax authority. We discussed RSS ontology design and provided a proof of concepts along with testing the ontology using RACER reasoner to check consistency and the ability of the ontology to query and logically classify RSS data and infer new concepts that are not explicitly asserted in the ontology.



- [15] Mark Pilgrim, "what is RSS", <http://www.xml.com/lpt/a/1080,2002>
- [16] The members of the RSSDEV Working Group, "RDF Site Summary (RSS) 1.0", <http://web.resource.org/rss/1.0/spec>
- [17] <http://www.protege.stanford.edu>
- [18] RDF schema, <http://www.purl.org/rss/1.0/schema.rdf>
- [19] Dublin core meta data element set: <http://www.dublincore.org/documents/dcse/>
- [20] OWL guide : <http://www.w3c.org/TR/owl-guide/>
- [21] Racer reasoner: <http://www.sts.tu-harburg.de/~r.f.moeller/racer/>
- [22] Egyptian Income tax authority: <http://www.incometax.gov.eg/map.asp>
- [22] M. Rahamatullah Khondoker, Paul Mueller, "Comparing Ontology Development Tools Based on an Online Survey", Proceedings of the World Congress on Engineering 2010
- [23] Sanjay Kumar Malik, Nupur Prakash and S.A.M Rizvi, "Developing an University Ontology in Education Domain using Protégé for Semantic Web", International Journal of Engineering Science and Technology, 2010

# Distributed Task Allocation in Multi-Agent System Based on Decision Support Module

Sally M. El-Ghamrawy  
Computers and Systems  
Department,  
Faculty of Engineering,  
Mansoura University, Egypt  
[Sally@mans.edu.eg](mailto:Sally@mans.edu.eg)

Ali I. El-Desouky  
Computers and Systems  
Department,  
Faculty of Engineering,  
Mansoura University, Egypt  
[ali\\_eldesouky@yahoo.com](mailto:ali_eldesouky@yahoo.com)

Ahmed I. Saleh  
Computers and Systems  
Department,  
Faculty of Engineering,  
Mansoura University, Egypt  
[aisaleh@mans.edu.eg](mailto:aisaleh@mans.edu.eg)

**Abstract-.** A Multi-Agent System (MAS) is a branch of distributed artificial intelligence, composed of a number of distributed and autonomous agents. In MAS, an effective coordination is essential for autonomous agents to reach their goals. Any decision based on a foundation of knowledge and reasoning can lead agents into successful cooperation, so to achieve the necessary degree of flexibility in coordination, an agent requires making decisions about when to coordinate and which coordination mechanism to use. The performance of any MAS depends directly with the right decisions that the agents made. Therefore the agents must have the ability of making right decisions. In this paper, we propose a decision support module in a distributed multi-agent system, which enables any agent to make decisions needed for Task allocation problem; we propose an algorithm for Task Allocation Decision Maker (TADM). Furthermore, a number of experiments were performed to validate the effectiveness of the proposed algorithm (TADM)); we compare the efficiency of our algorithms with recent frameworks. The preliminary results demonstrate the efficiency of our algorithms

**Keywords:** .Decision Making, Task allocation, Coordination Mechanism, Multi-Agent System (MAS)

## I. INTRODUCTION

The notion of distributed intelligent systems (DIS) [1] has been a subject of interest for number of years. A Multi-Agent system (MAS) is one of the main areas in the DIS. Any Multi-agent system consists of several agents capable of mutual interaction, with heterogeneous capabilities, that cooperate with each others to pursue some set of goals, or to complete a specific task. MAS used to solve problems which are difficult or impossible for an individual agent or monolithic system to solve. Agents are autonomous programs which can understand an environment, take actions depending upon the current status of the environment using its knowledge base and also learn so as to act in the future. In order to solve complex problems agents have to cooperate and exhibit some level of autonomy. Agents cooperate with each other to solve large and complex collaborative problems. Because the majority of work is completed through distributing the tasks among the cooperative agents, the decision support module is responsible on the most of the work. So the pros and cons in decision support module directly affect on the success of tasks completion and problem solving.

In [2], we proposed a Distributed Multi-Agent Intelligent System (DMAIS,) a general purpose agent framework, in which several interacting intelligent agents cooperate with some auxiliary agents to pursue some set of goals or tasks

that are beyond their individual capabilities. There are some modules must be considered in designing the Distributed Multi-Agent Intelligent System (DMAIS) that help in developing in the agent systems. There are many researches considered some modules when designing multi-agent systems (MAS) in different ways. Each autonomous agent in DMAIS must be able to decide how to behave in various situations, so in this paper our main concern is to propose an efficient decision support module that helps in the improvement of DMAIS performance.

Agents have attractive characteristics like: autonomy, reactivity, reasoning capability and social ability. These characteristics ensure that agent-based technologies are responsible for enhancing the decision support system capabilities beyond the capabilities of the old model. Any active decision support can be facilitated by the autonomy, reactivity and social ability of agents. Furthermore, the artificial view of agents can contribute towards stronger collaborative relationships between a human and a decision support system. These enormous interests of researchers in investigating the decision support in Multi-Agent Systems is due to the great benefits of combining the decision support technology and Agent-based technology with taking the advantage of the agent characteristics. In this sense, many researches [3-7] recognized the promise that agent -based technologies holds for enhancing DSS capabilities. Decision support module is a vital module in the success of DMAIS framework, due to the fact that any decision based on a foundation of knowledge and reasoning can lead agents into successful cooperation. So the performance of any multi-agent system depends directly with the right decisions that the agents made. Obviously, complex decision making tasks cannot be achieved by a single agent. Rather, it's achieved by efforts coordination of multiple agents possess different sets of expertise, attributes and assignment. This coordination among agents, which provide satisfactory solutions to problems among agents, needs many decisions that agents are required to make before this coordination can take place. So the Decision support module role is to allow agents to make the decision needed, which can help in the improvement of the DMAIS framework. In this paper, we extended the work done in DMAIS [2] by proposing a Decision support module in it, this module is concerned about taking the right decisions needed to allocate a specific task to a specific agent, by using the Task Allocation Decision Maker sub-module (TADM). So we spotted on these decisions due to the great importance for them in the improvement of agents' coordination in DMAIS framework. The rest of the paper is organized as following. Section 2 demonstrates the proposed

decision support module. Section 3 discusses the related work on task allocation in Multi-agent Systems. An algorithm for the Task Allocation Decision Maker (TADM) is proposed in section 4, showing the scenario of our algorithm in allocating tasks to specific agent/s. Section 5 shows the experimental evaluation and the results obtained after implementing the proposed algorithm. Section 6 summarizes major contribution of the paper and proposes the topics for future research.

## II. DECISION SUPPORT MODULE

In the DMAIS framework [2], the decision support module takes the information collected in the coordination and negotiation module, as shown in fig 1, so that it can help in the decision support process.

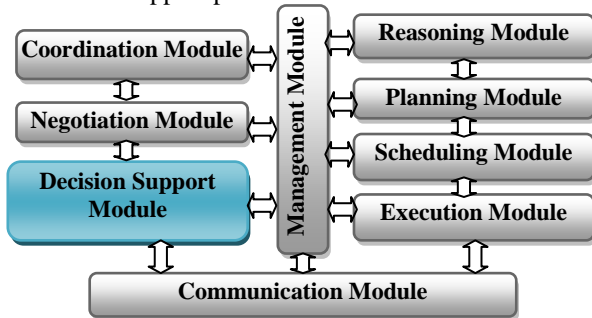


Figure 1. The Decision Support Module in DMAIS Framework

The Decision support module is activated when a decision is needed from the agent, we concerned about two main decisions that the agent may face: (1) the decision needed to allocate a specific task to a specific agent, making an effective decision for the task allocation problem is a critical job for any multi-agent system; it helps our DMAIS framework to complete its tasks and missions through cooperation among agents. The Task Allocation Decision Maker sub-module (TADM) is responsible for making this decision. (2) The decision needed to select appropriate coordination mechanism, when agents need to coordinate with another agent/s to accomplish a specific task.

The decision support module contains four main phases, as shown in figure 2, first is the decision knowledge management that contains:

- *The database*: contains the data that directly related to the decision problem (i.e. the performance measures, the values of nature states).
- *The knowledge base*: contains the descriptions for roles and structures of document and some knowledge of the problem itself, (i.e. Guide for how to select decision alternatives or how to interpret outputs).
- *The knowledge modeling*: is a repository contains the decision problem formal models and the algorithms and methodologies for developing outcomes from the formal models. Also contains different process models, each model is represented as a set of process and event objects.

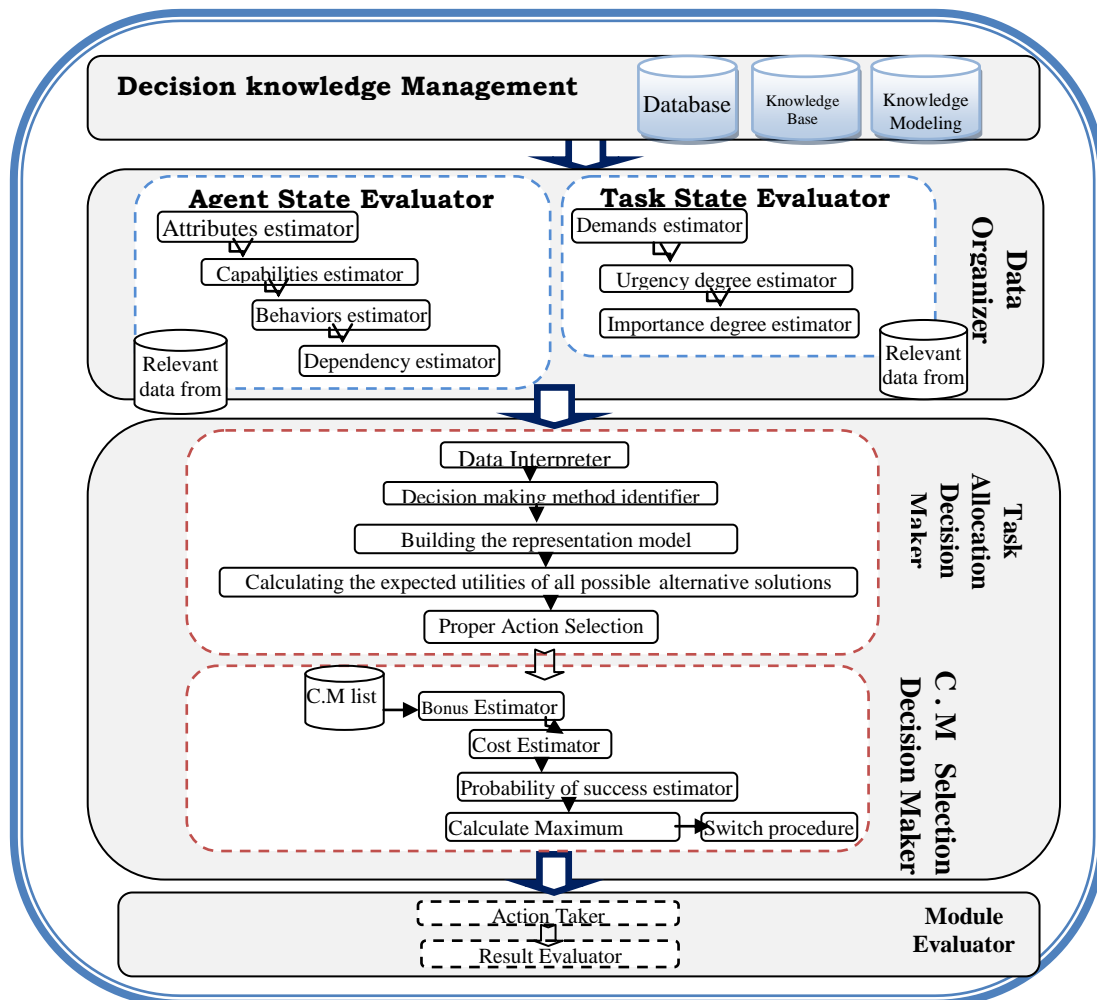


Figure 2. Architecture of the Decision Support Module

Second phase is the data organizer, which organizes agent's attributes using the agent state evaluator and organizes task's preferences by using task state evaluator. The agent state evaluator estimates the attributes and capabilities of agents, in our work we concerned on specific attributes and characteristics of the agent that will help in the decision making procedure, as shown in fig 3, where:

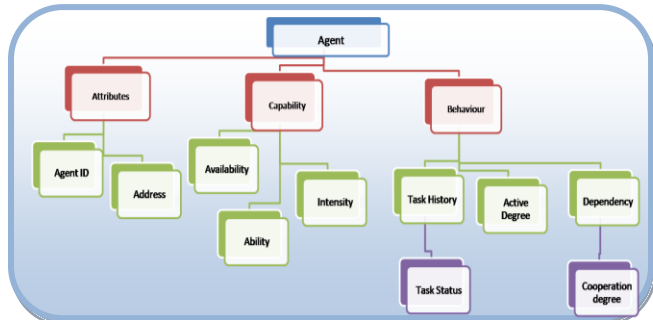


Figure 3. Attributes, Capabilities and behaviors of Agents

Agent attributes is defined as a tuple

$$\langle \text{AgentID}(A), \text{Addrs}(A) \rangle \quad (1)$$

Where  $\text{AgentID}(A)$  is the identity of the agent,  $\text{Addrs}(A)$  records the IP address of agent A. Agent Capability is defined as a tuple

$$\langle \text{Availability}(A), \text{Ability}(A), \text{Intensity}(A) \rangle \quad (2)$$

Where  $\text{Availability}(A)$  is the ratio of total number of successful agents, capable of accomplishing the task, to the total number of agents in the system,

$$\text{AVLB}(T) = \text{SUC}(A) / \text{TOT}(A) \quad (3)$$

$\text{Ability}(A)$  indicates number of agents capable of fulfilling the task.  $\text{Intensity}(A)$  refers to the number of tasks that the agents can accomplish per time unit.

Agent Behaviors is defined as a tuple

$$\langle \text{Task History}(A), \text{Active degree}(A), \text{Dependency}(A) \rangle \quad (4)$$

Where task history(A) stores the number of accomplished tasks that the agent(A) performed. Active degree(A) indicates the activity degree of a specific agent, this degree varies from agent to another according to many parameters (e.g. number of finished tasks, number of cooperation process). Agent Dependency(A) is concerned with the Cooperation Degree of an Agent(A) with respect to other agents in the system.

Third phase is divided into two main Sub-modules: (1) The Task Allocation Decision Maker (TADM), this sub-module will be discussed in details in the next sections, and (2) The Coordination Mechanism Selection Decision Maker (CMSDM), that takes the decision needed to select appropriate coordination mechanism, when agents need to coordinate with another agent/s to accomplish a specific task, it will be discussed in our future work. Both of them are responsible on making the proper decisions according to the input obtained from the Data organizer phase. The module evaluator phase is the fourth phase, it has two main goals: first, take the action that the third phase decided to be taken. Second, evaluate the results of taking this action.

### III. RELATED WORK

The task allocation in Multi-Agent Systems represent a problem which occupied to a large extends the researchers of decision support and artificial intelligence until our days.

Tasks allocation is defined, in [8], as the ability of agents to self-organize in groups of agents in order to perform one or more tasks which are impossible to perform individually. In our context, it is a problem of assigning responsibility and problem solving resources to an agent. There are two main benefits for Minimizing task interdependencies in the coordination process between the agents: First, improving the problem solving efficiency by decreasing communication overhead among the agents. Second, improving the chances for solution consistency by minimizing potential conflicts. The issue of task allocation was one of the earliest problems to be worked on in Distributed Artificial Intelligence (DAI) research. In this sense, several authors studied the problem related to Task Allocation especially in MAS. The researches in task allocation can be classified in to two main parts, centralized and distributed, based on utility/cost functions.

The researches that investigate the task allocation problem in a centralized manner: Zheng and Koenig [9] presented reaction functions for task allocation to cooperative agents. The objective is to find a solution with a small team cost and each target to be assigned to the exact number of different agents. This work assumed that there is a central planner to allocate tasks to agents. Kraus et al. [10] proposed an auction based protocol which enables agents to form coalitions with time constrains. This protocol assumed each agent knows the capabilities of all others, and one manager is responsible for allocating tasks to all coalitions. Pinedo [11] proposed a job shop scheduling treats the task allocation mostly in a centralized manner, and also ignores the communication cost. There are many drawbacks in the centralized task allocation like single point failure and bad scalability. To conquer these disadvantages, Task allocation in distributed environments has also been investigated. Davis and Smith [12] was the first in investigating a classic distributed task allocation in the multi-agent system using Contract Net Protocol (CNP), in which agents negotiate to assign tasks among themselves. Most of the subsequent literature on distributed task allocation is based on either contract net protocol or auctions [13]. The authors in [14] and [15] developed distributed algorithms with low communication complexity for forming coalitions in large-scale multi-agent systems. Abdallah and Lesser [16] provided a decision theoretic model in order to limit the interactions between agents and mediators. Mediators in this research mean the agents which receive the task and have connections to other agents. Mediators have to decompose the task into subtasks and negotiate with other agents to obtain commitments to execute these subtasks. However, their work concentrated on modelling the decision process of a single mediator. Sander et al. [17] presented a scalable and distributed task allocation protocol. The algorithm adopted in this protocol is based on computation geometry techniques but the prerequisite of this approach is that agents' and tasks' geographical positions are known. Weerd et al. [18] proposed a distributed task allocation protocol in social networks. This protocol only allows neighbouring agents to help with a task which might result in high probability of abandon of tasks when neighbours cannot offer sufficient resources. Dayong et al. [19] proposed an Efficient Task Allocation Protocol (ETAP) protocol based on the Contract Net approach, but more suitable for dealing with task allocation problems in P2P multi-agent systems with a



decentralized manner. It enables agents to allocate tasks not only to their neighbours but also to commit unfinished tasks to their neighbours for reallocation. In this way, the agents can have more opportunities to achieve solution of their tasks. Brahmi et al. [20] developed a decentralized and scalable method for complex task allocation for Massive Multi-Agent System, distributing the process of computing the optimal allocation among all agents based on the hypothesis: non conflict will be generated in the task allocation processes. Indeed, while being based on its Galois Sub-Hierarchy (GSH) and cooperation with other agents, each agent chooses the appropriate sub-task that ensures the global allocation optimality. Cheng and Wellman [21] used a market based protocol for distributed task allocation.

#### IV. TASK ALLOCATION DECISION MAKER (TADM)

The task allocation Decision Maker (TADM) main goal is to take the proper decisions of allocating tasks to the right agents. The first step in allocating tasks to specific agents is to take the decision whether the agents are capable of executing part or the entire task. The allocation decisions are made independently by each agent

The decisions needed for the Task allocation problem are mainly concerned about allocating the tasks to number of agents, whether these agents can complete its tasks by themselves or not. If agents can't achieve the task by themselves, they attempt to give a decision to specify other agents which have the appropriate capabilities and assign the task, or part of the task, to those agents. Fig 4 depicts the scenario of allocating tasks to specific agent/s in the TADM.

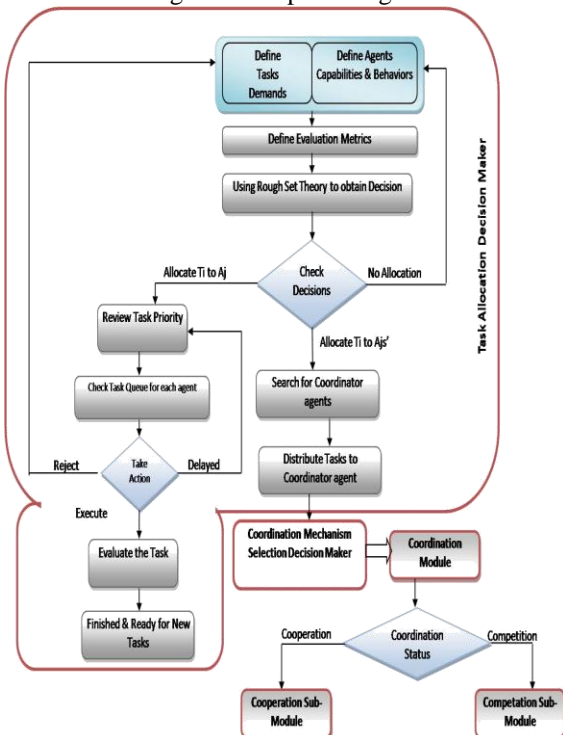


Figure 4: The Proposed algorithm used in TADM

First the capabilities and behaviors of agents must be defined, and also the demands of tasks, it's taken from the data organizer as an input to the TADM. The metrics that facilitate the decision making process must be evaluated, and then we

used the rough set theory for classifying these metrics to obtain the decision of allocating tasks to agent/s. Then this decision is tested, it may be one of three decisions: first, if the allocation cannot be done, in this situation the capabilities of agents and demands of tasks must be defined all over again, it may encounter any kind of error. second, if the decision is to allocate the task to only one agent, then task priority must be reviewed to check the task queue for each agent, and depend on this queue an action must be taken whether to delay, reject or execute this task. Third, if the decision is to allocate the task to more than one agent, then the coordinated agents must registered, and the sub-tasks must be distributed among those coordinated agents according to the algorithm in figure 5.

#### TADM algorithm

```
1. Agent Ai randomly select task T
2. AinCharge= Ai
3. For each a(i) in G
4.   Ai=Send request ( )
5.   Ai=Wait response ( )
6.   If time waited > expired time then
7.     Exit for
8.   End if
9.   Ai=Receive response ( )
10.  Ai=Process response ( )
11.  Ai=Store response ( )
12.  Max-value=0
13.  If helpfulness-value> Max-value then
14.    Max-value = helpfulness-value
15.    Nominated-agent= a(i)
16.  End if
17. Next
18. Xx:
19. Ai=Check (Nominated-agent ( ))
20. If helpfulness-value(Ai)< Nominated-agent then
21.   Ai= send Response(Nominated-agent)
22.   If reply (Nominated-agent) =1 then
23.     AinCharge= Nominated-agent
24.   Else
25.     Ai= find_snd_highst ( )
26.     Go to xx
27.   End if
28. Else
29.   AinCharge= Ai
30. End if
```

Figure 5: The Pseudo code for TADM algorithm

#### V. EXPERIMENTS EVALUATION

To evaluate the performance of TADM, we compare it with Efficient Task Allocation Protocol ETAP [18] and with the Greedy Distributed Allocation Protocol (GDAP) [19]. In order to validate the effectiveness of TADM algorithm and compare it with ETAP and GDAP two experiments are performed, as shown in figure 5; each experiment has its own goal and settings. In each experiment to evaluate the experiment results two metrics are evaluated the Efficiency Ratio and Run Time, which can be defined as follows:

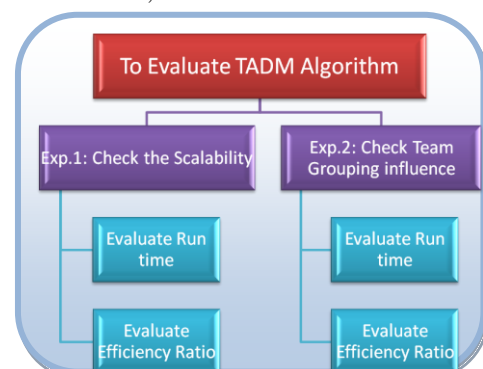


Figure 5: The experiments of Evaluating TADM algorithm

The Efficiency Ratio is the ratio between summation efficiency of finished tasks and the total efficiency expected of tasks. The efficiency of a task can be calculated as follows:

$$Eff_{(T)} = \frac{Reward_{(T)}}{Resource_{(T)}}$$

Where:

$Reward_{(T)}$  is the rewards gained from successfully finishing the task.  $Resource_{(T)}$  is the resources required for accomplishing the task.

Run Time is the time of performing TADM algorithm in the network under pre-defined settings. The unit of Run Time is millisecond. To investigate the effects of TADM, and compare it with ETAP and GDAP, a multi-agent system has been implemented to provide a testing platform. The whole system is implemented on a 6 Pc's with an Intel Pentium 4 processor at 300GHz, with 3GB of Ram, connected with network Ethernet 512Mbps. A network of cooperative agents is designed, in which most agent team are connected to each other, the generation of this network can follow the approach proposed in [22]. For each experiment, there are unified settings have to be specified.

Figure 6 shows the TADM when it's activated, and begins to perform its roles.

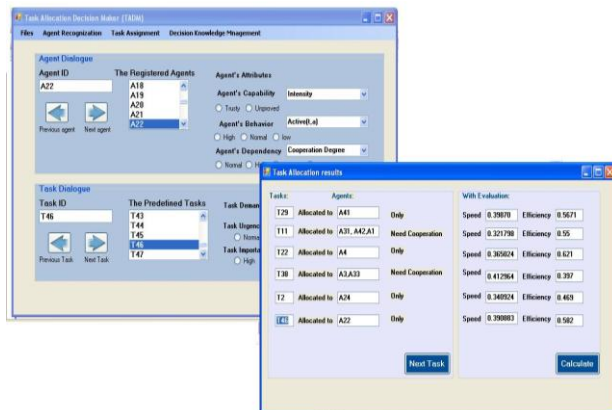


Figure 6: Screen shot of the TADM GUI

**In Experiment 1:** The main goal of this experiment is to demonstrate the scalability of TADM algorithm and compare it with ETAP and GDAP, as shown in figure 7 and figure 8, using same environment and settings, as follows:

- The average number of agent's team is fixed at 8.
- The number of agents range from 100 to 600, depending on the specific test.
- The number of tasks is range from 60 to 360.
- The number of resources types is range from 10 to 60.

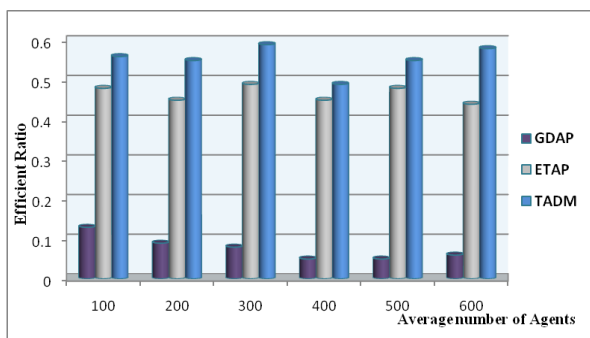


Figure 7: The Efficient Ratio on different number of agents

From figure 7, it is shown that when the number of agents is increasing, the Efficiency Ratio of TADM is much higher and more stable than of ETAP and GDAP that is continually descending with the increasing of agents.

Figure 8 shows the Run Time of TADM, GDAP and ETAP when the number of agents in the network change. It can be noticed that ETAP spends more time when there are more agents in the network, this is because ETAP make many reallocation steps which results in time and communication overhead rising. On the other hand, the time consumption of GDAP is steady during the entire test process and keeps a lower level than ETAP, this because GDAP relies on neighbouring agents only.

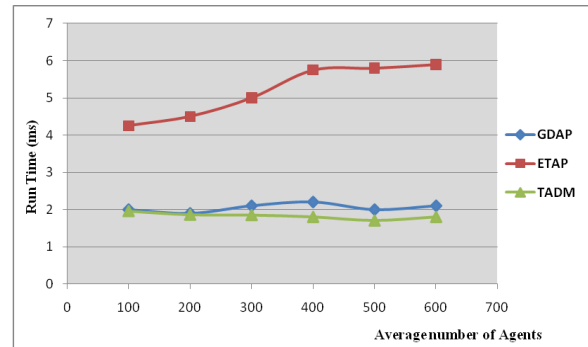


Figure 8: The Run Time on different number of agents

While TADM avoid these two drawbacks from the recent systems, and this experiment shows that the TADM performance is faster than that of GDAP and ETAP.

**In Experiment 2:** The main goal of this experiment is to test the influence of team grouping on TADM (i.e. show how different average number of agent's team influences the performance of TADM) and compare it with ETAP and GDAP, as shown in figure 9 and figure 10, using same environment and settings, as follows:

- The average number of agent's team is fixed at 8.
- The number of agents and tasks are 50 and 30 separately.
- The average number of resources for each type is 30 and the average number of resources required by each task is also 30.
- The number of resources types is 5.

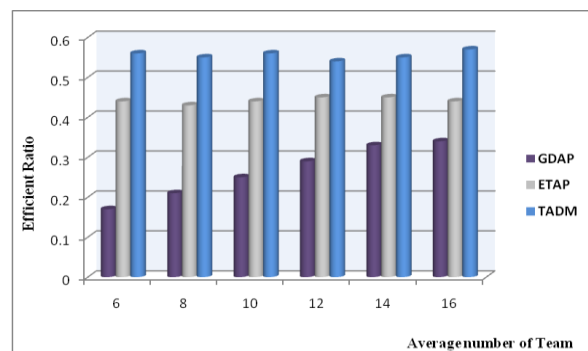


Figure 9: The Efficient Ratio on different number of agent's team

Figure 9 demonstrated that The Efficiency Ratio of TADM is much higher and more stable than that of GDAP and ETAP. GDAP performance is very low this is because task allocation in GDAP is only depending on neighbours of the agent. On the other hand, ETAP has better performance because it relies



not only on neighbours of the agent, but also other agents if needed.

The TADM has the higher performance and more stable than GDAP and ETAP, these results ensures and validates our algorithm.

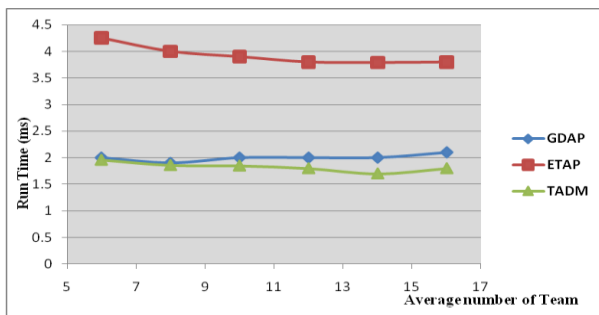


Figure 10: The Run Time on different number of agents' team

The Run Time of GDAP, ETAP, TADM in different number of agents' team is depicts figure 10. It's obviously that the Run Time of ETAP is higher than that of GDAP. As ETAP has to reallocate tasks when resources from neighbours are insufficient this lead to increase the reallocation steps and more time spending. While GDAP is steady due to its considering for only neighbours which could decrease the time and communication cost during task allocation process. But TADM takes the benefits of agent's team and also uses any other agent if needed, and in the same time reduces the steps taken to allocate the task which lead to decrease the time and communication cost during task allocation process. So this is why TADM has better Run Time that GDAP and ETAP.

## VI. CONCLUSION AND FUTURE WORK

In this paper, a decision support module is proposed which enables any agent to make decisions needed, focused on two main decisions: first, is the decision needed to allocate a specific task to a specific agent. Second, the decision needed to select appropriate coordination mechanism. And a survey of recent algorithms in task allocation in Multi-agent Systems is discussed. In addition an algorithm for the Task Allocation Decision Maker (TADM) is proposed showing the scenario of allocating tasks to specific agent/s. Finally, a preliminary experiment is then conducted, indicating that TADM has the scalability advantage comparing to most recent systems. We plan to propose new algorithm for the coordination mechanism selection in the decision support module. Also, we intend to propose a coordination module in our DMAIS framework.

## REFERENCES

- [1] Gruver, W., "Technologies and Applications of Distributed Intelligent Systems", IEEE MTT-Chapter Presentation, Waterloo, Canada, 2004.
- [2] A. El-Desouky and S. El-Ghamrawy. A Framework for Distributed Decision Support Multi-Agent Intelligent System. The 2008 International Conference on Artificial Intelligence ICAI08, WORLDCOMP'08 July14-17, 2008, Las Vegas, Nevada, USA.
- [3] T. Bui and J. Lee, "An agent-based framework for building decision support systems , " Decision Support Systems , vol. 25 , 1999 , pp. 225-237.
- [4] S.D. Pinson, J.A. Louca and P. Moraitis, "A distributed decision support system for strategic planning," Decision Support Systems , vol. 20 1, 1997, pp. 35-51.

- [5] Song Qiang, Lingxia Liu, "The Implement of Blackboard-Based Multi-agent Intelligent Decision Support System," iccea, vol. 1, pp.572-575, 2010 Second International Conference on Computer Engineering and Applications, 2010
- [6] Rustam and Bijan, "Pluralistic multi-agent decision support system: a framework and an empirical test, " Information & Management, vol. 41, Sept., 2004, pp. 883-898.
- [7] Wang Fuzhong, "A Decision Support System for Logistics Distribution Network Planning Based on Multi-agent Systems," dcabes, pp.214-218, 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, 2010
- [8] Totic, P.T., Agha, G.A.: Maximal Clique Based Distributed Group Formation For Task Allocation in Large-Scale Multi-Agent Systems. In: Ishida, T., Gasser, L., Nakashima, H. (eds.) MMAS 2005. LNCS (LNAI), vol. 3446, pp. 104-120. Springer, Heidelberg (2005).
- [9] X. Zheng and S. Koenig. Reaction functions for task allocation to cooperative agents. In Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), pages 559-566, Estoril, Portugal, May 2008
- [10] S. Kraus, O. Shehory, and G. Taase. Coalition formation with uncertain heterogeneous information. In Proceedings of 2th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003), pages 1-8, Melbourne, Australia, July 2003.
- [11] Pinedo, M., "Scheduling: Theory, Algorithms and Systems", Prentice Hall, 1995.
- [12] Davis, R., and Smith, R., "Neogitation as a metaphor for distributed problem solving," Artificial Intelligence, 20, 63-100, 1983.
- [13] R. McAfee, and J. McMillan, "Auctions and bidding," Journal of Economic Literature, 25, 699-738, 1987.
- [14] K. Lerman and O. Shehory. Coalition formation for largescale electronic markets. In Proceedings of 4th International Conference on Multi-Agent Systems, pages 167-174, Boston, Massachusetts, USA, July 2000. IEEE Computer Society.
- [15] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. Artificial Intelligence, 101(1-2):165-200, May 1998
- [16] S. Abdallah and V. Lesser. Modeling task allocation using a decision theoretic model. In Proceedings of 4th Autonomus Agents and Multiagent Systems (AAMAS 2005), pages 719-726, Utrecht, Netherlands, July 2005.
- [17] P. V. Sander, D. Peleshchuk, and B. J. Grosz. A scalable, distributed algorithm for efficient task allocation. In Proceedings of 1st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2002), pages 1191-1198, Bologna, Italy, July 2002.
- [18] M. D. Weerd, Y. Zhang, and T. Klos. Distributed task allocation in social networks. In Proceedings of 6th Autonomus Agents and Multiagent Systems (AAMAS 2007), pages 500-507, Honolulu, Hawaii, USA, May 2007.
- [19] Dayong Ye, Quan Bai, Minjie Zhang, Khin Than Win, Zhiqi Shen, "An Efficient Task Allocation Protocol for P2P Multi-agent Systems," ispa, pp.11-18, 2009 IEEE International Symposium on Parallel and Distributed Processing with Applications, 2009.
- [20] Zaki Zaki Brahmi, Mohamed Mohsen Gammoudi, Malek Ghenima: Cooperative Agents Based-Decentralized and Scalable Complex Task Allocation Approach Pro Massive Multi-Agents System. ACIIDS (2) 2010: 420-430
- [21] 4.Cheng, J., andWellman, M. , "TheWALRAS algorithm: A convergent distributed implementation of general equilibrium outcomes," Computational Economics, 12, 1-24 ,1998.
- [22] D. J. Watts and S. H. Strogatz. Collective dynamics of small world' networks. Nature, 393:440-442, 1998

# Multiplayer Enhanced Make Square Game in the Net

Md. Ekramul Hamid  
Department of network Engineering  
College of Computer Science  
King Khalid University  
Abha, Kingdom of Saudi Arabia  
e-mail:ekram\_hamid@yahoo.com

Sudip Saha  
Software Engineer  
Protocom Technology  
Sydney, NSW-2000  
Australia.

Mohammed Qayyum  
Department of Computer Engineering  
College of Computer Science  
King Khalid University  
Abha, Kingdom of Saudi Arabia

**Abstract**— In this work, the authors attempt to create a successful Java socket program to implement the “Make Square” game in the net. The game is very popular among the children. Even though this game is for kids, this one can also be played by an adult because it can be quite tricky, require concentration and a little bit of intelligence. The goal in this game is to make more small squares. A player will win the game, if he can complete maximum number of squares after the game is over. Here client/server technology is used to implement socket programming. Since the game is implemented by java, so it is platform independent and portable. Many players in many different groups can play the game on the net. To make the game more interesting we enhance its feature by adding hidden lines. This makes the game more attractive and challenging. The Java features like Networking, Graphics, Layout Management, Package and Interface, Exception Handling, I/O, Applets, AWT Controls and Event handling etc. [2-4] are used to create the game. The Make Square game consists of more than 1700 lines of code in 12 classes. Five of these classes are part of the server side and rest seven is part of the client side. The Make Square game is running properly in a network.

**Keywords**-component; Make Square, node, socket programming, AWT, 3D object, GUI, client/server technology

## I. INTRODUCTION

There are so many games like the “Make Square” that are really silly and funny. But not all of them are like this that require concentration and a little bit of intelligence. These are the puzzle games. The controls for these games are very simple and usually need to use only the mouse. Player can click on different objects on the screen to trigger all sorts of actions. Here in this paper we propose the development of online games using Java. We use Applets because everyone has an Internet browser, so everyone is able to play this game without installing the JDK!

Java is a multiplatform, object-oriented programming language created by Sun Microsystems. Creating games on Java is simple than creating games on languages such as C++ due to the advantages of the Swing library that comes as part of the Java API.

The “Make Square” is a very popular game played on paper by drawing a board layout as figure 1. Here, the authors attempt to implement tele-playing version of the game in the network [1]. To play the game, each player connects to server using client program. It does not need any third party server software to play the game.

The Make Square playing board consists of nodes, where nodes are arranged as a m x n matrix. A typical board of a

Make Square is shown in the figure 1. Basically, the nodes in the board are made of filled circles.

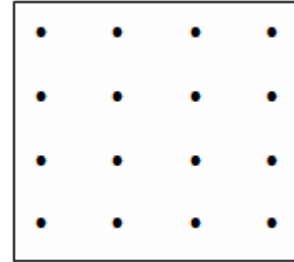


Figure 1. Playing Board (m x n Matrix of nodes)

## A. General Rules of the Game

A player connects two nodes with a line in his turn. Once he puts the line, his turn is over, and next players turn is start and so on. In this way, the game goes on in a round robin fashion until all the squares are completed. A player will win the game, if he can complete maximum number of squares after the game is over.

Let two players A and B are playing the game. In The first turn, player A puts his first move A(1,1) and B puts his first move B(1,1) as shown in figure 2. Here, each player's turn represented by PLAYER (i,j), where i and j are turn and move of PLAYER respectively.

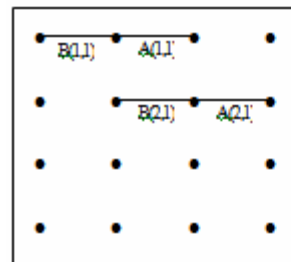


Figure 2. Turns of A and B

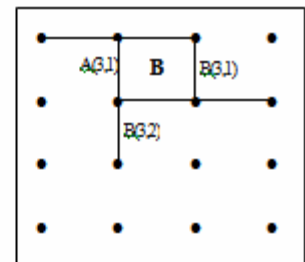


Figure 3. B wins a square

The player who puts the last line to form a complete square — wins the square. Figure 3 shows B wins a square. When no more square is remaining, the game is over and the player who completes more squares is the winner. To make the game more interesting we enhance its feature by adding hidden lines. This makes the game more attractive and challenging.

### B. Enhanced Rules of the Game

**Rule1:** One player (say A) can hide one line, which is not visible to others. He cannot put another hidden line until his hidden line is disclosed to all.

**Rule2:** A player (let A) can put a line, there are two possible places,

#### 1. Where no line exists

If this line makes a complete square with hidden line(s), he wins the square. Since he (A) disclosed the hidden line(s), he will get another chance to put a new line. Figure 4 shows A and B put hidden lines (indicated by dashed line) and figure 5 shows A disclosed the hidden lines in the next move and wins the square.

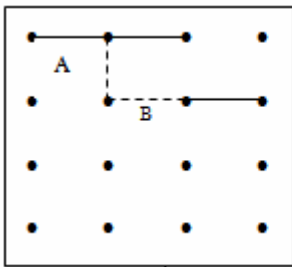


Figure 4. Hidden lines

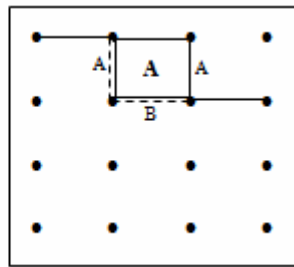


Figure 5. Disclosed hidden line

#### 2. On another existing hidden line, which is not visible to him (A), then the hidden line will disclosed only to him (A) and get a chance to put a new line.

**Rule3:** A player (let A) can put a hidden line in two possible places

#### 1. Where no line exists, then it will visible to that player (A) only and invisible to others.

#### 2. On another existing hidden line, which is not visible to him (A), then this line will also consider as a hidden line. In this case a player can't win the square even though that hidden line makes a complete square.

## II. GAME SETUP

The Make Square game setup is very simple because of its strong Graphical User Interface (GUI). One player can setup the game either in solo mode or multi-player mode.

### A. Solo Mode

One player can play Make Square in Solo mode. Here his opponent may be an intelligent agent. He has to enter his name, choose his pen color and game size (number of nodes in the row and column) and start the game.

### B. Multi-Player Mode

In this mode, many players can play the game. Here some players (max. 4) together make a group where one player is leader and others are members. Each player has to enter his name, pen color and choose status (as a leader/member). If he joins as group leader, he has to select game size. Otherwise he has to choose group number to play (Figure 6). This

information is passed to the server and the server broadcast the update player list. The user then see a list of all available players and any one from the group can start the game by clicking the start button.

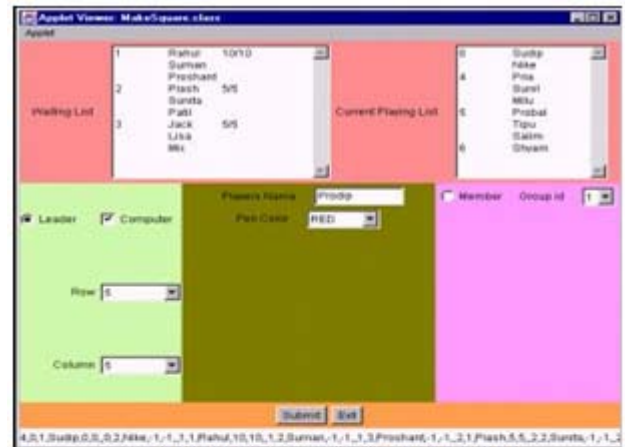


Figure 6 Game setup environments in multi-mode

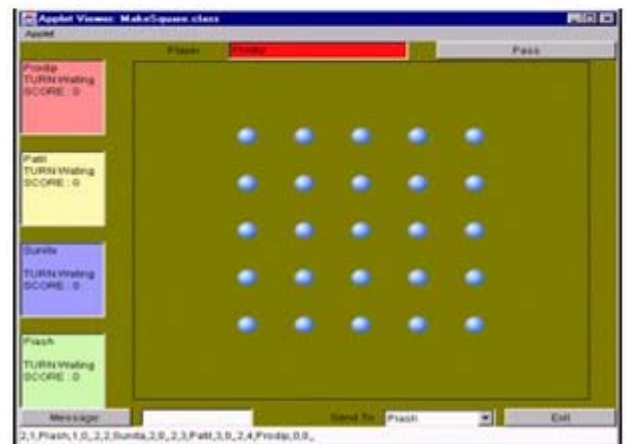


Figure 7. Game playing environment in multi-mode

### C. Different Buttons and Windows

#### Buttons:

- start** : allows to start the games
- message** : allows to send messages to others
- exit** : quitting the game
- pass** : to skip a turn

#### Windows:

**Waiting list window:** shows the player list in the queue.

**Score Board window:** shows player name and score of that player. The player, who wins a square, gets 1 point.

**Message window:** at any time during the play, players may converse using this window.

The game-playing environment is shown in the figure 7.

#### D. Features of the Game

- Easy game setup
- Attractive board layout
- Hidden line option
- Chatting facility
- Descriptive score board
- Multi-group support
- Durability (against system crash)

### III. ROLE OF CLIENT/SERVER

The client-server computing is a distributed application structure that partitions tasks between the providers of a resource, called servers, and service requesters, called clients. Often clients and servers communicate over a computer network on separate hardware, but both client and server may reside in the same system. A server machine is a host that is running one or more server programs which share their resources with clients. A client does not share any of its resources, but requests a server's content or service function. Clients therefore initiate communication sessions with servers which await incoming requests. Figure 8 shows the client/server communication.

#### A. Role of Client

- Client provides GUI as per the requirements of the game.
- Client facilitates the player to start a new game by invoking start option.
- Client shows the players name, group id in a list.
- Client updates each player, screens and sends correct messages.

#### B. Role of Server

- Server provides multiple client connections so that many players can play.
- Server continuously listens to the different requests from the multiple clients.
- Server maintains all the clients information.
- Server maintains the play sequence and notifies the player about his turn

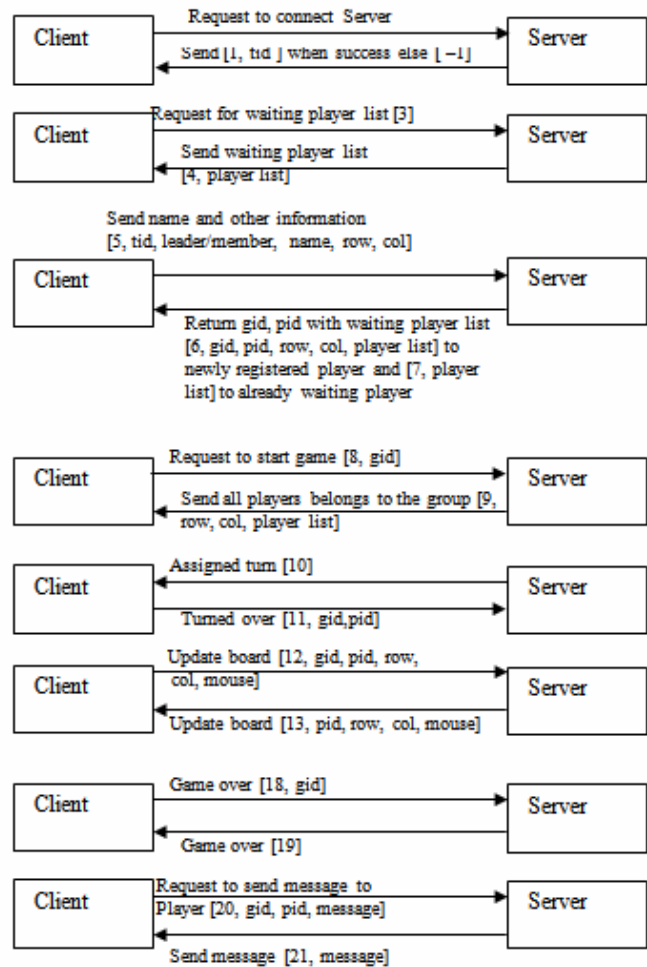


Figure8. Client Server Communication

### IV. DATA STRUCTURES USED

#### A. Server Side

##### Data Structure for Player Handling

The data structure for player handling in the server side is shown in the figure 9.

When a client initially connects to server, it assigns a temporary id (tid) to the client until he becomes a member of the game. In figure 9, socket indicates the socket address from which the client connects. After getting the membership, each players information such as group id (gid), players id (pid), socket number (socket), players name (name), number of rows and columns (row, col) in the playing board will be stored in the following structure as shown in the figure 10.



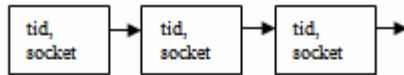


Figure 9. Temporary linked list (tlist)

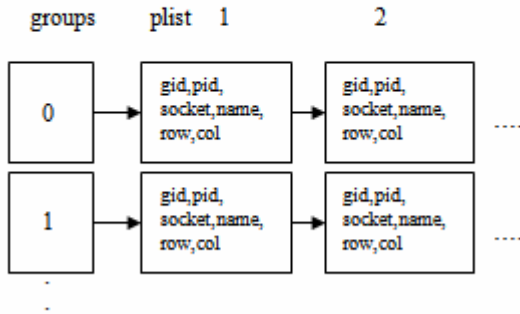


Figure 10. Player lists in different groups

- Server maintains an array of linked list for different playing groups, each group have one linked list [6].
- First node in the linked list contain the leader information remaining nodes contain members information.
- In the array of linked list there may have two types of group, waiting group and playing group.
- Server differentiates between playing group and waiting group using the row in the leader node. If it is -1, implies the group is playing.
- When a group's game is over, then its corresponding list will be free.

### B. Client Side

#### Data Structure for Player Handling

The data structure for player handling in the client side is shown in the figure 11. In the figure 11, each client holds all players information such as group id (gid), player id (pid), player name (name), Score, hidden line, pen.

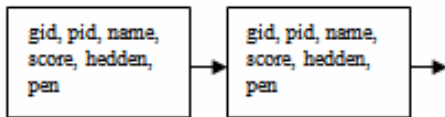


Figure 11. A linked list for player handling

#### Data Structure for the Playing Board

Figure 12 represents the playing board structure, which is a 2 X 2 matrix. In the matrix each element (other than nodes) store an integer value. The elements  $(i, 2j+1)$  [where  $i=1,3,5,\dots$  and  $j=0,1,2,3,\dots$ ] store the CENTER value and others store the TLBR (Top/Left/Bottom/Right) value.

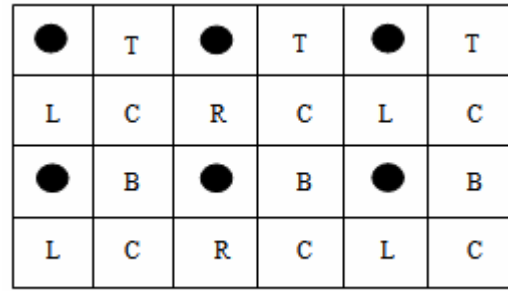


Figure 12. Playing board structure

The center (C) bit structure in Figure 13 is maintained in the center cell of each square. The bit positions 1,2, and 4 in the center integer that represents the total lines around the center. Bit no 1 represents One (1) Line, Bit no 2 represents Two(2) Lines and Bit no 4 represents Three(3) Lines around the center. Remaining Bit positions 64, 32, 16 and 8 represent Line positions. If Bit position 1, 2, 4 contain 1 then Remaining Bit position 64, 32, 16 and 8 represents the Player id.

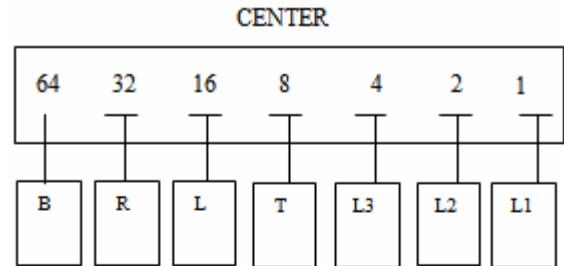


Figure 13. Bit structure for the center (C) value

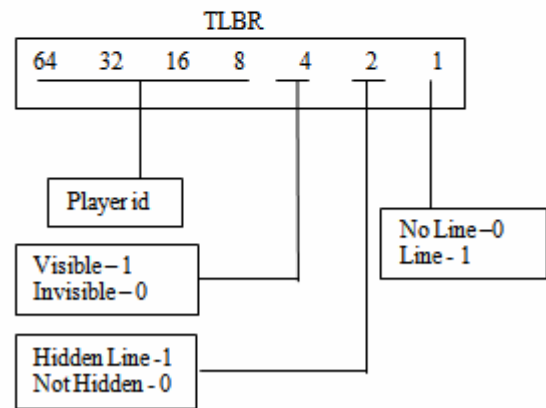


Figure 14. Bit structure for the TLBR value

In the TLBR bit structure in figure 14, bit position 1 represents the Line status (1 for existence). Bit positions 2 represents whether the line is a hidden or not. Bit position 3 represents whether the Line is Visible or not. Remaining Bit position 64, 32, 16 and 8 represents Player id that put the line.

## V. THE GAME CODE

The Make Square game consists of more than 1700 lines of code in 12 classes [1-4]. Five of these classes are part of the server side and rest seven is part of the client side. Most of the classes and its functions are given in appendix -I.

## VI. ACTIVITY DIAGRAM

Activity diagram are use to model the dynamic aspects of a system. With the help of this diagram one can show the flow of an object as it moves from state to state at different points in a system [7].

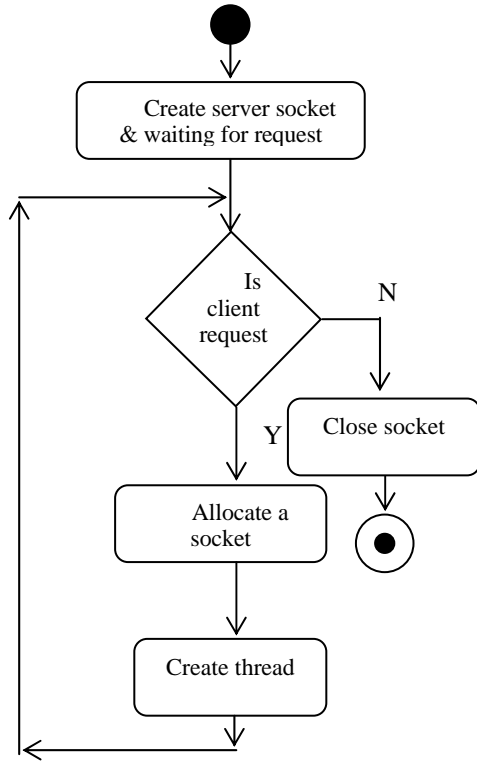


Figure 15. Activity diagram (server side)

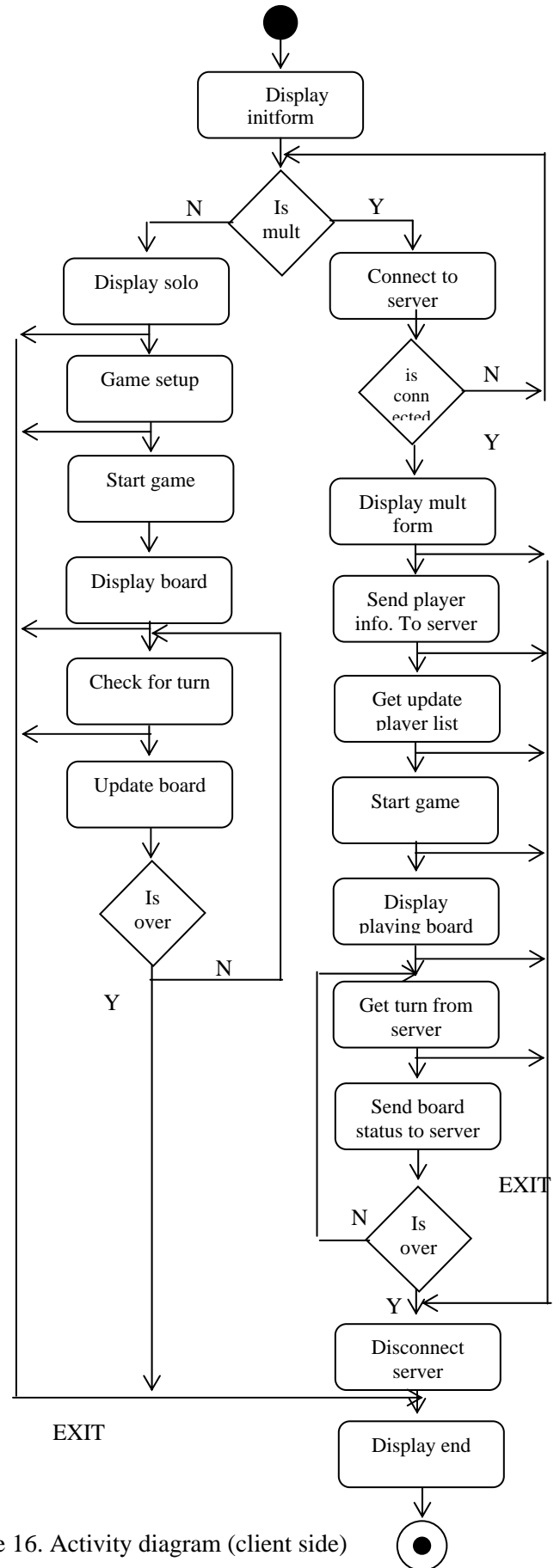


Figure 16. Activity diagram (client side)



## VII. RESULTS AND DISCUSSIONS

The Make Square game is running properly in a network. The playing board updates real-time. The game can survive from system crash, media failure and power failure of clients. The Make Square game is very popular to us because it is quite easy to play. The game is intended for entertainment purpose and net version of it makes it more interesting. It can be market as a commercial product like other games.

## VIII. CONCLUSION

In this Project work, we have created an applet represents a complete client/server, multi-player board game "Make Square" using Java programming. Almost all the important java concepts are implemented here. The motivating objective is to develop a basic understanding of java programming in the network environment.

In the future the following improvements to the software are recommended:

- More efficient GUI
- Make computer logic more intelligent as a player

## REFERENCES

- [1] M. E. Hamid and S. Saha, "Make Square", MCS 2<sup>nd</sup> semester project, Department. of Computer Science, Fergusson College, University of Pune, India, June 2001.
- [2] H. Schildt, *The Complete Reference*, Tata McGraw-Hill Publishing Company Limited, New Delhi, 2001.
- [3] E. Bruce, *Thinking in Java*, Prentice Hall PTR, 1999.
- [4] S. C. Horsmann, *Core Java*, Volume-I&II, Sun MicroSystem Inc. 2000.
- [5] E. Balagurusamy, *Programming in Java*, Tata McGraw-Hill Publishing Limited, New Delhi, 1995.
- [6] L. Yedidyah, J. Moshe, M. T. Aaron, *Data Structures using C and C++*, Prentice-Hall of India Private Limited, New Delhi 110001, 1998.
- [7] B. Grady. R. James, *The Unified Modeling Language User Guide*, Person Education Asia, New Delhi 2001.

## APPENDICES

### APPENDIX-I

#### 1) **class CONNECTtoSERVER extends Thread**

This class connects client to server and creates a thread for continually listening the server information using following functions:

- **public int ConnectToServer()**

It connects client to server. If successful, return a message, otherwise -1

- **public void run()**

It continuously listens server information and call message handler.

#### 2) **public class MakeSquare extends Applet**

This class makes user interface.

- **public void init()**

init function displays a screen where user choice modes either solo or multi.

- **public void mousePressed (MouseEvent me)**

When a mouse button is clicked this function call select line method of the board class.

- **public void MessageHandeler()**

This method handles the message sent by server.

- **public void BoardMng (int r,int c,int iMouseBtn,int Pid)**

It manages the playing board. Whether player turn is active or his turn is over and accordingly informs the server.

- **public void actionPerformed (ActionEvent ae)**

When a particular button is clicked this function takes appropriate action.

- **private void MultForm()**

This form provides users information to the server.

- **private void OpenForm()**

It shows a player the update list of players and update the waiting player list.

- **private void BoardForm()**

It shows playing board.

- **public void UpdateScoreBoard(int pid, boolean Bright, int score)**

It updates the score board and other information's (playing or waiting and hidden status)

- **public Color SelectPenColor(int pen, boolean Bright)**

It selects a pen color according to pen provided by argument (pen).

#### 3) **class INITform**

- **public void ShowInitForm()**

This function show the initial form of the applet.

#### 4) **class SOLOfrom**

- **ShowSoloFrom()**

This function shows a form in solo mode.

#### 5) **class PCData**

This class is used for storing information of players in client side. Informations are player id, group id, pen, score, hidden, name,

- **interface CONSTANTList**

Here all constants are declared.

6) **class BOARD implements CONSTANTSlist**

- **public void DisplayBoard()**

This function displays the board of the game.

- **public void DrawLine(int r, int c, int Pid)**

It draws a line according to row, column and player id (using pid's pen color).

- **public void UpdateMatrix(int r, int c, int iMouseBtn, int Pid)**

It updates matrix (board) according to the row, column, mouse, and pid.

- **public void ModifyCenter(int r, int c, byte TLRBid, int Pid)**

- **public void ModifyCenter(int r, int c, int iMouseBtn, int Pid)**

Modify the center bit structure.

- **public void ModifyTLRB(int r, int c, byte TLRBid, int Pid)**

It updates top, left, right, bottom bit structure.

- **public void SelectLine(MouseEvent me)**

It selects line according to mouse click. left button for solid line and right button for hidden line.

7) **class Object3D**

- **Object3D(int RL, int GL, int BL, Applet app)**

This function shows 3D balls according to color provided by its argument.

- **public void Draw3DBall(Graphics gc, int x, int y, int r)**

Draw a ball according to the arguments (radius, xy position)

8) **public class Server5**

This class creates a server socket according to specified address and waiting for client.

9) **class PSData**

This class is used for storing information of players in server side. Information's are player id, group id, socket, name, row, and column.

10) **class Tdata**

It stores temporary information of players like temporary id, and socket.

11) **class CLIENTmng**

- **synchronized public void AddClient()**

This is used to add a new client and provide a temporary id to the user.

- **public void BroadCast(int cata, int gid, int pid, String msg)**

It is used to broadcast message to client. Its category parameter implies:

0 → to all client connected to server.

1 → to all waiting and playing player.

2 → to all waiting except a particular player.

3 → to all player in a group.

4 → to all player except a particular player.

5 → to all waiting player.

6 → to all waiting player except a particular player.

10 → to a particular player.

- **synchronized public int ProcessClientMsg()**

To process clients messages.

12) **class CREATEclient extends Thread**

This class creates an instance that listens a particular client information.

## AUTHORS PROFILE



**Md. Ekramul Hamid** received his B.Sc and M.Sc degree from the Department of Applied Physics and Electronics, Rajshahi University, Bangladesh. After that he obtained the Masters of Computer Science degree from Pune University, India. He received his PhD degree from Shizuoka University, Japan. During 1997-2000, he was a lecturer in the Department of Computer Science and Technology, Rajshahi University. Since 2007, he has been serving as an associate professor in the same department. He is currently working as an assistant professor in the college of computer science at King Khalid University, Abha, KSA. His research interests include Software development, speech enhancement, and speech signal processing.



**Sudip Saha** received his B.C.S and M.C.S degree from Pune University, India. He received Post Graduation Diploma in Information Technology from University of New England, Australia. He is currently working as a software engineer in Protocom Technology, Sydney, Australia. His research interests include telecommunication and mobile technologies.



**Mohammed Qayyum**, Currently working as Lecturer in the Department of Computer Engineering at King Khalid University, Kingdom of Saudi Arabia. Before joining at KKU, he worked as a Assistant Professor under the Department of Computer Science and Engineering at MJCET, Hyderabad, India. Prior to that, He also worked as a Systems Engineer with designation as Member Technical Staff for Ikanos Communications Inc. Bangalore, India. He completed his Master of Technology degree in Software Engineering from JNT University, Hyderabad, India. He is keen and desperate for research in Data Replication Techniques of Mobile Adhoc Networks with Database issues.

# Descriptive System For Creating Awareness In The Electoral Process In Nigeria Using Information Technology

B. A. Onyekwelu

Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Osun State  
[bukolajc@yahoo.co.uk](mailto:bukolajc@yahoo.co.uk)

&

Dr. D.T. Akomolafe, MBCS

Department of Computer Science, Joseph Ayo Babalola University, Ikeji- Arakeji, Osun State, Nigeria  
[dtakomolafe@yahoo.com](mailto:dtakomolafe@yahoo.com)

Corresponding author email address is [dtakomolafe@yahoo.com](mailto:dtakomolafe@yahoo.com)

## ABSTRACT

Knowledge is power, as is popularly said, and lack of knowledge of the electoral process of one's nation makes one a subject, rather than citizen. What makes the difference between citizens and subjects is the type and volume of information possessed. This paper discusses the electoral process in Nigeria in relation to the principal actors in the process, namely, the electorates, the political players, the electoral body, the Judiciary and the Government. Their roles of each principal actor are highlighted. The current state of awareness of the

electoral process in Nigeria is explained as well as factors leading to this state. Information Technology and its growth in Nigeria are reviewed. The Methodology for creating people's awareness towards the electoral process in Nigeria is proposed and evaluated. The challenges facing the advancement of Information Technology in the country are enumerated and a conclusion is drawn.

**Keywords:** electoral process, information, Nigeria, Government, Technology.

## 1. INTRODUCTION

The popular saying, "knowledge is power", is not an exaggeration. One of the major factors that affect the electoral process, and almost every other process in Nigeria and other developing countries is the lack of knowledge of the process by the majority of the participants. One important way to impart knowledge is by creating awareness. Majority of the people in developing countries, Nigeria in particular are not aware of the details of the electoral process. It is important for the entire populace to be made aware that the electoral process involves more than just coming out to vote. The ability to vote correctly is determined by proper awareness of all the other key events in the electoral process. Proper awareness also promotes active participation and feeling of citizenship among the populace, which is crucial to the success of any developing nation (Weldeab, 2010)

Unlike in time past, the advent of Information Technology (IT) and its rapid spread in Nigeria has made the work of awareness creation a surmountable task. IT can be applied to most, if not all the key events of the electoral process in such a way that as each event is unfolding, the people are equipped with proper awareness. These will eventually make an average person

in the process has correct information and proper understanding of each event thereby setting a high level of transparency in the electoral process, which will in turn bring about a higher level of trust in the governance of the nation.

## 2. THE ELECTORAL PROCESS IN NIGERIA

The electoral process in Nigeria began immediately the country gained independence in 1960. Being a newly independent country, the government was made up of a coalition of different political parties, which were the Nigerian People's Congress (NPC), the National Council of Nigeria and the Cameroons (NCNC) led by Nnamdi Azikiwe, who became Nigeria's maiden Governor-General in 1960. The liberal Action Group (AG) formed the opposition. The nation seceded from its British legacy in 1963 by declaring itself a Federal Republic. Elections were held in 1965, amid dubious electoral circumstances. (Wikipedia, 2010). Since then, the nation has gone through several political eras.

The Nigeria Electoral system is the single member constituency type with competitive

multi-party and the first past the post winner system. The method of voting used in five out of six past elections, that is, in 1979, 1983, 1999, 2003 and 2007 was the Open Ballot System (OBS). The OBS entails a procedure in which the prospective voter goes through a process of accreditation, receives a ballot paper from the appropriate poll official and thereafter makes the confidential thumb impression in favour of the political party or candidate of choice in a secret voting compartment before dropping the ballot in the box positioned in the open, in the full glare of officials, security agents and party agents. (Okop, 2006)

There are five key groups involved in the electoral process and these are the electorates, the political players, the electoral officials, the Judiciary and the Government. Each of these groups has their distinct roles in the electoral process.

### **2.1. The Electorates**

These are the citizens, and they make up the general populace; the people of the land. In any democracy, there is a clear difference between citizens and subjects. The subjects passively allow the Government to initiate and carry out public policies, while the citizens actively participate in the rituals of democracy (Lyons and Alexander, 2000). The difference between the two is determined by their level of awareness. This group usually takes active part in two major events in the electoral process, which are voters' registration and voting. In Nigeria, as in most other nations, the voting age is 18 years and above. This group constitutes the most vital component of the electoral system and they determine the success or failure of the system. Consequently, they are supposed to be made aware and sensitized of every development in the electoral process.

### **2.2. The Political Players**

These are the main players in the electoral process. They include the political aspirants as well as everyone who is registered as a member of a political party. Some of the key events that involve this group, apart from the registration and voting, are:

- Registration of parties
- Registration of members into each party
- Registration of candidates

- Establishment of strategies of winning elections
- Primary Elections
- Elections
- Proper monitoring
- Presentation of candidates for elective positions

### **2.3. The Electoral Body**

This group is made up of the electoral commission set up by the government to conduct elections and personnel employed either as a full or ad hoc staff. These are the personnel employed specifically to implement the various key events that make up the electoral process as contained in the laws that set up the body. They can be described as the bridge between the electorates and the political players. They are involved in most of the activities of the electoral process. The events that involve the electoral officers, as described by (Nelson, 2001), are:

- Registration of voters or updating of registration records;
- Parties and candidates' registration;
- educating voters;
- accrediting observers and monitors;
- establishing an electoral campaign period;
- oversight of the process and its machinery;
- preparing for, and then holding the vote and count;
- announcing the results; and
- inauguration of the newly elected officials

Nigeria as a country has a population of about 120 million people, out of which, there are about 60 million registered and eligible voters spread across 120,000 polling centres. As a result of this, election supervision and manning of the centres require about 500,000 officials, a greater number of which are temporary or ad-hoc staff, recruited and trained, usually very late in the elections process (Okop, 2006).

The officials are usually appointed and coordinated under the umbrella of the electoral body which is put in place by the government. This has always brought the electoral body under fire from many quarters, such as political parties and civil society any time there is a failure in the electoral process. They have accused the

commission of being susceptible to pressure from the ruling Party (Owen, 2003).

## **2.4 JUDICIARY**

The judiciary is charged with its normal responsibility of dispensing justice in the electoral process. Any dispute, complaint relating to the election as it affects any of the players and intra party disputes are adjudicated upon by the judiciary

## **2.5. The Government**

This includes the ruling party, as well as the political office holders who might belong either to the ruling party or an opposition party. The Nigerian Government has always played an active role in the electoral process, and particularly in the putting in place the electoral body. In 1992, the ruling Government went as far as creating two political parties, and creating the guidelines and manifesto guiding each of the parties.

## **3. THE PRESENT STATE OF AWARENESS**

Presently, the state of awareness of the electoral process in Nigeria is very low, especially among the first group of participants, the electorates. This is due to several factors, some of which are;

- Lack of proper education of the electorate on the electoral processes, due to lack of adequate funding to the appropriate authorities. .
- Lack of proper media access as a result of unbalanced reporting and unequal advertising rates. Government owned print and electronic media give preferences to the ruling party in the provision of access above other parties and candidates.
- Irregular Power supply, which in turn limits the electorates' access to electronic media.
- Poverty, which makes an average man on the street want to focus his attention on earning his daily bread rather than making effort to be educated on the electoral process of the nation
- Bad Precedence, as a large percentage of the electorates is disillusioned because of the past experiences they had in the electoral process in the country.

All these factors must be taken into consideration in other to design an effective People Awareness program for the electoral process of Nigeria.

## **4. INFORMATION AND INFORMATION TECHNOLOGY**

Information Technology (IT) can be defined as computer-based activities that are derived from the convergent fields of micro-electronics, computing, and tele-communications and that have led to the reorganization of the processes of production, distribution, and circulation in society. (Nwachuku, 1994). It is the combination of computer, electronics and media to inform impart and educate for the attainment of corporate goals. It is the major tool that is used to generate and disseminate information. It is a known fact that information is an important resource that may be generated, shared, and utilized in decision-making. (Graham, 1980) defined information as data that has been processed. The importance of information cannot be overstressed; it is a major resource that determines the success or failure of any organization. Lack of information creates a huge gap in any given society, while availability of Information and access to knowledge make very big impact on progress of all societies. Effective use of Information Technology helps to reduce the knowledge gap. (Akomolafe & Eludire, 2009) opined that information is an important ingredient used in decision making for the attainment of aims, aspiration, goals and objectives and (Akinyokun, 1999) identified three forms on information that can be generated and disseminated using IT and these are text, video and sound. (Falaki, 2004) clearly explained how each of these forms could be generated and disseminated using IT.

In the last 15 years, Information Technology has created a huge impact on world society. (Bhalesain, 2007). IT has enormous potential as a tool not only for improving governance, but also to enhance the standard of living of the people. (Nair and Prasad, 2002).

### **4.1. The Growth of Information Technology in Nigeria**

All around the world, Governments and people are beginning to truly appreciate the ability of IT to stimulate rapid development in all sectors of the economy (Ajayi, 2003), and Nigeria is no exception. The Federal Government began an



Information Communication Technology (ICT) revolution in 1999, when she took the following steps:

- 1) Approval of policies for the major sectors of the industry: National Telecommunications Policy, National Information Technology Policy, National Space Policy and National Biotechnology Policy.
- 2) Liberalization of the sector
- 3) Right priority status given to ICT

In 2001, the Federal Executive Council (FEC), Nigeria, approved the National Information Technology Policy, and established the National Information Technology Development Agency (NITDA) to implement the policy (Ajayi, 2003). NITDA has mounted several projects, some of which are The Public Service Network (PSNet), Human Capacity Building and Mobile Internet Unit. However, the major challenge that NITDA has faced is that of inadequate funding.

The conducive environment created by the Federal Government and the liberalization of the sector propelled a significant growth in the use of ICT as shown by Goshit (2004) and summarized in Table 1:

**Table 1: Growth rate of ICT Infrastructures in Nigeria, between 2000 and 2004**

	2000	2001	2002	2003	2004
Number Of Connected Mobile Lines	-	1.6M	2.05M	3.1M	3.8M
NUMBER OF ISPS	18	30	30	35	35

## 5. THE DESCRIPTIVE METHODOLOGY

The electoral process is a long and complex process that relies heavily on information. The gathering, processing and dissemination of information between and within the key players in the electoral process are better facilitated and done by IT. IT has inbuilt capabilities, tools and functions that are useful in the processing, storing, retrieval and communicating information appropriately. These facilities and their relevance and usefulness to the electoral process are described hereafter

1. Provision of Information technology infrastructures: This is done first providing funds for the provision of hardware and software infrastructures, as well as internet services, and ensuring that they are put in place. It is highly important that the electorates have access to the

facilities. Though the use of mobile lines has a wide spread in Nigeria, the use of personal computers has not received the same response. The Mobile lines were initially beyond the reach of the masses at the beginning, but this has changed. The nation needs to experience this also with Personal Computers. It is worthy of note that a good number of Nigerian youths today use mobile phones with Internet facilities, but this is only common among the youths, and rare among the older generation of electorates.

2. Web and Portals: A portal is essentially a gateway and it is a website that serves as a starting point to access a variety of information and resources. It is imperative to ensure that the key players in the electoral process, the political parties, the candidates, and particularly the electoral commission make their presence known on the internet. The Federal Government of Nigeria accepted the decision to apply advanced technology in the electoral process in 1999, and since then, the electoral commission has began efforts to establish an effective presence on the internet for the dissemination of information on the Commission's activities, including the transmission of results (Guobadia, 2005)

3. Digitization of materials relevant to the electoral process and creation of information repository. An e-Library can be created to store texts and e-books relating to the electoral process online.

4. Effective search engine customized to meet the need of an average information seeker concerning the electoral process, and the state of the nation in general.

5. Audio and Video on demand: It is possible to generate these information using IT and to store, retrieve, and communicate them. It is a known fact that most most people would rather watch and listen, rather than read. Hence, Audio and Video online can be developed by these players to make their activities clear to each other and attract more people to information, rather than text. Conferences, workshops, rallies can be recorded and clips made available online.

6. The key players can also use the popular FaceBook, SMS, and Email etc to make their activities known to the public and for intra and inter exchange of information. Also interested members of the public including the players can

react to any of the activities by using these facilities.

7. Creating more awareness for effective use of IT: The most important point is propagation among masses. This can be done through the use of seminars and workshops, IT trainings, Media advertisements and Bill boards, and a host of others. Just as in some States of the Federation where Government compelled all the staff to go on computer training, the same thing can be done all over the nation, and in all Government Agencies. The people must be made to understand the gains of Information Technology and the fact that its use will make the electoral process clearer and thus, more open.

## 6. CHALLENGES

The challenges of using IT to create awareness in the electoral process are:

- a) Inadequate funding. This challenge is both societal and private, in the sense that, not only is the Government inadequately funding moves geared towards the use of IT, but also, the standard of living of the average Nigerian makes it difficult for them to give priority to Information Technology.
- b) Few number of IT specialists. Just like in some other professions in Nigeria, there are lots of 'specialists', but very few who are actually well-trained and who know what they are doing. A good number of computer personnel, 'engineers' as they are called in a lot of places, actually practice trial by error, just like the auto-mechanics and some other artisans.
- c) The extremely weak public power supply. This is the main paralyzing challenge to the development of IT in any sector in Nigeria. Only those who are willing to spare the extra fund to provide alternative means of power supply can truly enjoy IT facilities in the Country.
- d) The misuse of Information Technology by a few miscreants in the nation is another major challenge. Even though the security bodies like EFCC and ICPC are trying, the IT crime, particularly cyber crime is on the increase in the nation.
- e) The Percentage of women who are IT users, or even IT aware, is less, compared to men. This gap needs to be reduced for IT to make relevant impact on creating awareness towards the nation's electoral process.

## 7. CONCLUSION

So far, Nigeria has done fairly well in adopting Information Technology, but the pace of development has been affected by the economic situation of the country, as well as the unavailability of necessary infrastructures to support Information Technology. Globally, the cost of computer hardware has reduced drastically, making it more available to the common man. Also, the use of Mobile Phone technology to connect to the internet makes it a useful tool in creating awareness across the country. Just as Information Technology has been fully integrated in many sectors of the economy, particularly industry, banking and other financial institutions, in business and in the educational sector, it is hoped that it will also receive full integration in the Government and in the nation's electoral process as time goes on.

It is highly imperative that the challenges facing the proper utilization of Information Technology in creating awareness in the country be addressed, in order to move forward. There is need for adequate funding of the agencies involved, like the electoral commission where electoral process is concerned. IT specialists also need to be well trained and mobilized, for effective training. The ominous problem of irregular power supply needs to also be addressed, as it is the major hindrance to the effective use and spread of information technology in the country. The issues of e-security needs to be intently looked into, and incorporated strongly into every security bodies and law enforcement agencies in Nigeria, to keep abreast the growing wave of cyber crimes in the nation.

Finally, all these efforts must be backed up with massive mobilization of the people, the nation's electorates, to become IT literates.

## REFERENCES

- Ajayi, G.O. (2003). "NITDA and ICT in Nigeria, *Round Table on Developing Countries Access to Scientific Knowledge*", *The Abdus Salam ICTP, Trieste, Italy*
- Akinyokun, O.C. (1999) "Principles and Practise of Computing Technology", International Publishers Limited, Ibadan
- Akomolafe, D.T and Eludire, A.A. (2008), "The Role of Information and Communication Technology in National Development". *Journal of School of Science*, Adeyemi College of Education, Ondo, Ondo State, Nigeria. Vol.1 No 1 Pages 102-107
- Bhalesain, P. (2007). "Effective use of Information Technology (IT) for propagation of Buddhism", *IIT Roorkee*, [www.yba.org.in](http://www.yba.org.in)???
- Falaki, S.O. (2002) "Information Technology in Nigeria: Now or Never" (FUTA) Inaugural Lecture Series 29) Unix Computer Services Limited, Akure.
- Goshit, T. (2004). "Nigeria's Need for ICT", *SP.259 Technology and Policy in Africa*, <http://ocw.mit.edu/NR/rdonlyres/Special-Programs/SP-259Spring-2006/891209EE-E63B-4617-BA9D-7635A63C754B/0/goshit.pdf>
- Graham C. Lester (1980) "Data Processing; Polytech Publishers Limited, Stockport.
- Guobadia, A. (2005). "FEDERAL REPUBLIC OF NIGERIA – ELECTION ADMINISTRATION, IMPROVING THE QUALITY OF ELECTION MANAGEMENT", *A conference of Commonwealth Chief Election Officers*, organised by the Commonwealth Secretariat in co-operation with the Election Commission of India The Ashok Hotel, New Delhi, 24-26 February 2005
- Lyons, W. and Alexander, R. (2000). "A Tale of Two Electorates: Generational Replacement and the Decline of Voting in Presidential Elections". *The Journal of Politics*, Vol. 62, No. 4 (Nov., 2000), pp. 1014-1034 Published by: Cambridge University Press on behalf of the Southern Political Science Association Stable URL: <http://www.jstor.org/stable/2647863>
- Nair, K and Prasad, P. (2002). "Development through Information Technology in developing countries: Experiences from an Indian State". *The Electronic Journal of Information Systems in Developing Countries*. (2002), 8, 2, 1-13, [www.ejisd.org](http://www.ejisd.org).
- Nwachuku, M. (1994). "Development of Information Technology in Nigeria. *In Information Technology in some selected countries*". © The United Nations University, 1994 <http://www.unu.edu/unupress/unupbooks/uu19ie/uu19ie0d.htm>
- Okop, U. (2006). "A PAPER ON THE VOTING SYSTEM IN NIGERIA", PRESENTED AT THE Association of Electoral Administrators SEMINAR HELD IN BLACKPOOL, ENGLAND IN FEBRUARY, 2006 . [http://www.aea-elections.co.uk/annualseminars/download/s/nigeria\\_text.pdf](http://www.aea-elections.co.uk/annualseminars/download/s/nigeria_text.pdf)
- Owen, O. (2003). "The Nigerian Elections". © *Guardian News & Media* 2008 Published: 4/11/2003 <http://www.buzzle.com/editorials/4-11-2003-38918.asp>
- Weldeab, R. (2010): "Engaging Youth in Decision-Making: A Path toward Active Youth Citizenship in Eritrea", *Economic In Reform Feature Service, Publ Centre for International Private Practice*. <http://www.cipe.org/publications/fs/pdf/011510.pdf>. Assessed on March 26, 2010.
- Wikipedia. (2010) : History of Nigeria website

## ENCHANCING AND DERIVING ACTIONABLE KNOWLEDGE FROM DECISION TREES

1. P.Senthil Vadivu  
HEAD, Department Of Computer Applications,  
Hindusthan College Of Arts and Science,  
Coimbatore-641028. Tamil Nadu, India.  
Email: sowju\_sashi@rediffmail.com

2.Dr.(Mrs) Vasantha Kalyani David.  
Associate Professor,  
Department of Computer Science,  
Avinashilingam Deemed University  
Coimbatore, Tamil Nadu, India  
Email: vasanthadavid@yahoo.com

### Abstract

*Data mining algorithms are used to discover customer models for distribution information, Using customer profiles in customer relationship management (CRM), it has been used in pointing out the customers who are loyal and who are attritors but they require human experts for discovering knowledge manually.*

*Many post processing technique have been introduced that do not suggest action to increase the objective function such as profit. In this paper, a novel algorithm is proposed that suggest actions to change the customer from the undesired status to the desired one. These algorithms can discover cost effective actions to transform customer from undesirable classes to desirable ones. Many tests have been conducted and experimental results have been analyzed in this paper*

**Key words :** CRM,BSP,ACO, decision trees, attrition

### 1. Introduction

Researchers are done in data mining. Various models likes Bayesian models, decision trees, support vector machines and association rules have been applied to various industrial applications such as customer relationship management,(CRM)[1][2] which maximizes the profit and reduces the costs, relying on post processing techniques such as visualization and interestingness ranking.

Because of massive industry deregulation across the world each customer is facing an ever growing number of choices in telecommunication industry and financial services [3] [10].The result is that an increasing number of customers are switching from one service provider to another. This Phenomenon is called customer “churning “or “attrition”.

A main approach in the data mining area is to rank the customers according to the estimated likelihood and they way they respond to direct

marketing actions and compare the rankings using a lift chart or the area under curve measure from the ROC curve. Ensemble based methods are examined under the cost sensitive learning frameworks. For Example, integrated boosting algorithms with cost considerations.

A class of reinforcement learning problems and associated techniques are used to learn about how to make sequential decisions based on delayed reinforcement so as to maximize cumulative rewards.

A common problem in current application of data mining in intelligent CRM is that people tend to focus on, and be satisfied with building up the models and interpreting them, but not to use them to get profit explicitly. More specifically, most data mining algorithms only aim at constructing customer profiles; predict the characteristics of customers of certain classes. Example of this class is: what kind of customers are likely attritors and kind are loyal customers?

This can be done in the telecommunications industry. For example, by reducing the monthly rates or increasing the service level for valuable customers.

Unlike distributional knowledge, to consider actionable knowledge one must take into account resource constraint such as direct mailing and sales promotion [14]. To make a decision one must take into account the cost as well as the benefit of actions to the enterprise.

This paper is presented with many algorithms for the Creation of decision tree, BSP (Bounded Segmentation Problem), Greedy-BSP and ACO (Ant Colony Optimization) which helps us to obtain actions for maximizing the profit and finding out the number of customer who are likely to be loyal.

## 2. Extracting Actions in Decision Tree

For CRM applications, using a set of examples a decision tree can be built, which is described by set of attributes such as name, sex, birthday, etc., and financial information such as yearly income and family information such as lifestyles and number of children.

Decision tree is used in vast area of data mining because one can easily convert methods into rules and also obtain characteristics of customers those who belong to a certain class. The algorithm that is used in this paper do not only rely on prediction but also it can classify the customers who are loyal and such rules can be easily derived from decision trees.

The first step is to extract rules when there is no restrictions in the number of rules that can be produced. This is called as unlimited resource case [3]. The overall process of the algorithm is described as follows:

### Algorithm 1:

Step 1: Import customer data with data collection, data cleaning, data preprocessing and so on.

Step 2: A Decision tree can be build using decision tree learning algorithm[11] to predict, if a customer is in desire status or not. One improvement for the decision building is to use the area under the curve of the ROC curve [7].

Step 3: Search for optimal actions for each customer using the key component proactive solution [3].

Step 4: Produce reports, for domain experts to review the actions that deploy the actions.

### 2.1 A search for a leaf node in the unlimited resources

This algorithm search for optimal actions and transforms each leaf node to another node in the more desirable fashion. Once the customer profile is built, the customers who are there in the training examples falls into a particular leaf node in a more desirable status thus the probability gain can then be converted into expected gross profit.

When a customer is moved from one leaf to another node there are some attribute values of the customer that must be changed. When an attribute value is transformed from V1 to V2, it corresponds to an action that incurs cost which is defined in a cost matrix.

The leaf node search algorithm searches all leafs in the tree so that for every leaf node ,a best destination leaf node is found to move the customer to the collection of moves are required to maximize the net profit.

The domain specific cost matrix for the net profit of an action can be defined as follows:

$$P_{Net} = P_E * P_{gain} - \sum_i COST_{ij} \quad (1)$$

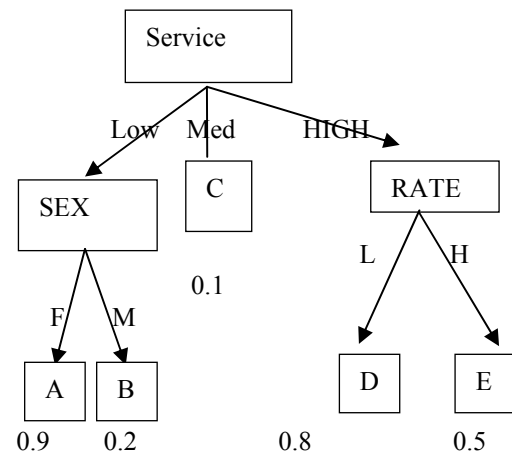
Where  $P_{Net}$  denotes the net profit,  $P_E$  denotes the total profit the customer in the desired status,  $P_{gain}$  denotes the probability gain, and  $COST_{ij}$  denotes the cost of each action involved.

The leaf node search algorithm for searching the best actions can be described as follows:

### Algorithm: leaf-node search

1. For each customer x, do
2. Let S be the source leaf node in which x falls into;
3. Let D be a destination leaf node for x the maximum net profit  $P_{Net}$ .
4. Output (S, D,  $P_{Net}$ );

An example of customer profile:-



Consider the above decision tree, the tree has five nodes. A, B, C, D, E each with the probability of customers being a loyal. The probability of attritors simply “1” minus this probability.

Consider a customer Alexander who’s record states that the service=Low (service level is low), sex=M (male), and Rate =L (mortgage rate is low). The customer is classified by the decision tree. It can be seen that Alexander falls into the leaf node B, which predicts that Alexander will have only a 20 percent chance of being loyal. The algorithm will now search through all other leafs (A, C, D & E) in the decision tree to see if Alexander can be “replaced” into a best leaf with the highest net profit.

1. Consider the leaf node A. which do not have a high probability of being loyal(90%), because the cost of action would be very high if Alexander should be

changed to female). So the net profit is a negative infinity.

2. Consider leaf node C, it has a lower probability of being loyal, so we can easily skip it.

3. Consider leaf node D the probability gain is 60 percent (80 percent - 20 percent) if Alexander falls into D, the action needed is to change service from L (low) to H (high).

4. Consider leaf E, the probability gain is 30 percent (50 percent - 20 percent), which transfers to \$300 of the expected gross profit. Assume that the cost of the actions (change service from L to H and change rate from L to H). Is \$250, then the net profit of the jack from B to E is \$ 50 (300-250).

Clearly, the node with the maximum net profit for Alexander is D, that suggest actions of changing the service from L to H.

### 3. COST MATRIX:-

Each attribute value changes incur cost and the cost for each attribute is determined by the domain experts. The values of many attributes, such as sex, address, number of children cannot be changed with any reasonable amount of money. These attributes are called "hard attributes". The users must assign large number to every entry in the cost matrix.

Some values can be easily changed with reasonable costs, these attributes such as the service level, interest rate and promotion packages are called "soft attributes".

The hard attributes should be included the tree building process in the first place to prevent customers from being moved to other leafs is because that many hard attributes are important accurate probability estimation of the leaves.

For continuous value attributes, such as interest rate which is varied within a certain range. the numerical ranges be discretized first for feature transformation.

## 4. THE LIMITED RESOURCE CASE: POSTPROCESSING DECISION TREES:-

### 4.1 BSP (Bounded Segmentation Problem)

In the previous example that is considered above each leaf node of the decision tree is a separate customer group. For each customer group we have to design actions to increase the net profit. But in practice the company may be limited in its resources. But when such limitations occur it is difficult to merge all the nodes into K segments. So to each segment a responsible manager can apply several actions to increase the overall profit.

Step 1: Here, a decision tree is build with collection  $S(m)$  source leaf nodes and collection  $D(m)$  destination leaf nodes.

Step 2: Consider a constant,  $k.(K < m)$ , where  $m$  is total number of source leaf nodes.

Step 3: Build a cost matrix with attributes  $U$  and  $V$ .

Step 4: Build a unit benefit vector, when a customer belongs to positive class

Step 5: Build a set of test cases.

The goal is to find a solution with maximum net profit. by transforming customers that belongs to a source node  $S$  to the destination node  $D$  via, a number of attribute value changing actions.

### GOALS:

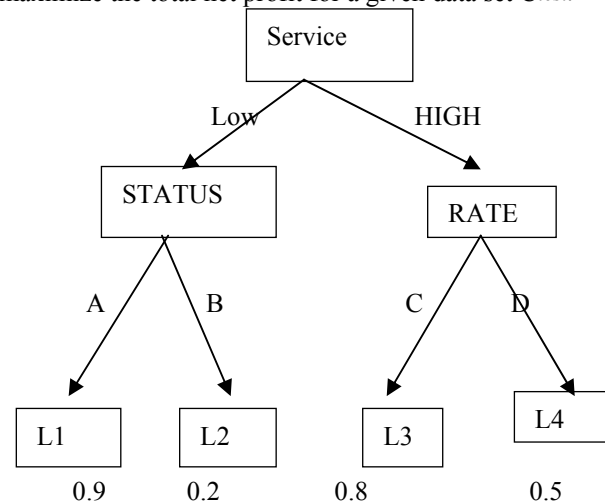
The goal is to transform a set of leaf node  $S$  to a destination leaf node  $D$ ,  $S \rightarrow D$ .

### ACTIONS:

In order to change one has to apply one attribute value changing action. This is denoted by

$\{Attr, u \rightarrow v\}$ .

Thus the BSP problem is to find the best  $K$  groups of source leaf nodes  $\{Group\ i=1, 2, \dots, k\}$  and their corresponding goals and associated action sets to maximize the total net profit for a given data set  $C_{test}$ .



Example: To illustrate the limited resources problem, consider again our decision tree in above figure. Suppose that we wish to find a single customer segment  $\{k=1\}$ . A candidate group is  $\{L2, L4\}$ , with a selected action set  $\{service \leftarrow H, Rate \leftarrow C\}$  which can transform the group to node  $L3$ . assume that group to leaf node  $L3$ ,  $L2$  changes the service level only and thus, has a profit gain of  $(0.8-0.2)*1-0.1=0.5$  and  $L4$  has a profit gain of  $(0.8-0.5)*1-0.1=0.2$ . Thus, the net benefit for this group is  $0.2+0.5=0.7$ .

As an example of the profit matrix computation, a part of the profit matrix corresponding to the source leaf node.  $L2$  is as shown in table, where  $Aset1=\{status=A\}$ .  $Aset2=\{service=H, Rate =C\}$  and  $Aset3=\{service=H, Rate=D\}$ . here, for convenience we ignore the source value of the attributes which is dependent on the actual test cases.



TABLE1 An example of the profit matrix computation

Aset1(L2)(Goal = ->L1)	Aset2(L2)(Goal = ->L3)	Aset3(L2)(Goal = ->L4)
0.6	0.4	0.1
...	...	...

TABLE 2 : lustrating the Greedy-BSP algorithm

Source nodes	Aset1 (goal= ->D1)	Aset2 (goal= ->D2)	Aset3 (goal= ->D3)	Aset4 (goal= ->D4)
S1	2	0	1	1
S2	0	1	0	0
S3	0	1	0	0
S4	0	1	0	0
Column sum	2	3	1	1
Selected actions		X		

Then the BSP problem becomes one of picking the best k columns of matrix M such that the sum of the maximum net profit value for each source leaf node among the K columns is maximized. When all  $P_{ij}$  elements are of unit cost, this is essentially a maximum coverage problem, which aims at finding K sets such that the total weight of elements covered is maximized, where the weight of element is the same for all the sets. A special case of the BSP problem is equivalent to the maximum coverage problem with unit costs. Our aim will then to find approximation solutions to the BSP problem.

#### Algorithm for BSP:

Step 1: Choose any combination of k action sets  
Step 2: Group the leaf nodes into k groups  
Step 3: Evaluate the net benefit of the action sets on the group  
Step 4: Return the k action set with associated leaf node

Since the BSP needs to examine every combination of k action sets, the computation complexity is more.

To avoid this we have develop the Greedy algorithm which can reduce the computational cost and guarantee the quality of solution.

We consider the intuition of the Greedy BSP algorithm using an example profit matrix M as shown in table. Where we assume a k=2 limit. In this table each number a profit  $P_{ij}$  value computed from the

input parameters. Greedy BSP algorithm processes this matrix in a sequenced manner for k iterations. In each iteration, it considered adding one additional column of the M matrix, until it considered all k columns.

Greedy BSP algorithm considers how to expand the customer group by one. To do this, it considers which addition column will increase the total net profit to a highest value we can include one more column.

#### 5. IMPROVING THE ROBUSTNESS USING MULTIPLE TREES:

The advantage of the Greedy BSP algorithm is that it can significantly reduce the computational cost while guaranteeing the high quality of the solution at the same time. In Greedy BSP algorithm, built decision tree always choose the most informative attribute as the root node.

Therefore, we have also proposed an algorithm referred to as Greedy BSP multiple which is based on integrating an ensemble of decision trees in this paper [16],[5],and [15].

The basic idea is to construct multiple decision trees using different top ranked attributes as their root nodes. For each set of test cases, the ensemble decision trees return the median net profit and the corresponding leaf nodes and action sets as the final solution.

Thus, we expect that when the training data are unstable, the ensemble based decision tree methods can perform much more stable as compared to results from the single decision tree.

Algorithm Greedy BSP Multiple:

Step 1: Given a training data set described by P attributes

1.1 Calculate gain ratios to rank all the attributes in a descending order.

1.2 For i=1 to p

Use the ith attribute as the root node to construct the ith decision tree

End for

Step 2: take a set of testing examples as input

2.1 For i= 1 to p

Use the ith decision tree to calculate the net profit by calling algorithms Greedy BSP

End for

2.2 return k actions sets corresponding to the median net profit.

Since Greedy BSP multiple relies on building, multiple decision trees to calculate the median net profit different sampling can only affect the construction of a small portion of decision trees. Therefore. Greedy BSP Multiple can produce net profit less variance.

## 6. AACO (ADAPTIVE ANT COLONY OPTIMISATION):

The searching process of ACO is based on the positive feedback reinforcement [4][12]. Thus, the escape from the local optima is more difficult than the other Meta heuristics, therefore, the recognition of searching status and the escape technique from the local optima are important to improve the search performance of ACO. Regarding the recognition of searching status, the proposed algorithm utilizes a transition of the distance of the best tour. A period of which each ant builds a tour represents one generation.

Regarding the recognition of searching status, the proposed utilizes a transition of the distance of the best two (shortest 2). A period of which each ant builds a tour represents one generation.

There is a cranky ant which selects a path let us not being selected and which is shortest.

### ADAPTIVE ANT COLONY ALGORITHM

1. Initialized the parameters  $\infty$ , t max, t, sx, sy
2. for each agent do
3. Place agent at randomly selected site an grid
4. End for
5. While (not termination)  
// such at  $t \leq t_{max}$
6. for each agent do
7. Compute agent's Fitness  $f(\text{agent})$   
And activate probability  $P_a(\text{agent})$   
According to (4) and (7)
8.  $r \leftarrow \text{random}([0,1])$
9. If  $r \leq P_a$  then
10. Activate agent and move to random  
Selected neighbor's site not  
Occupied by other agents
11. Else
12. Stay at current site and sleep
13. end if
14. end for
15. adapting update parameters  $\infty$ ,  $t \leftarrow t+1$
16. End while
17. Output location of agents

### 7. DIFFERENCE FROM A PREVIOUS WORK

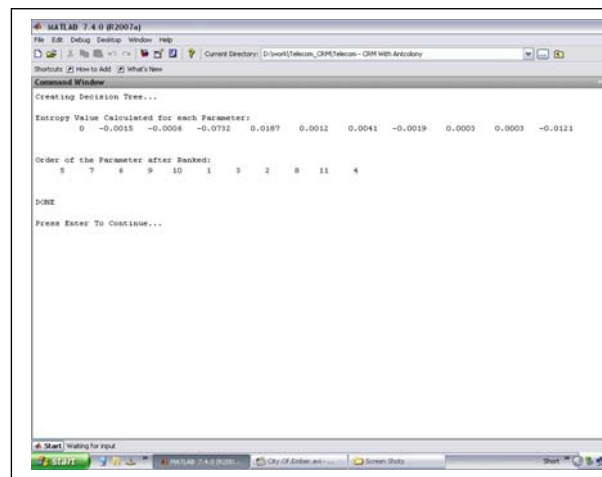
Machine learning and data mining research has been distributed to business practices by addressing some issues in marketing. One issue is that the difficult market is cost sensitive. So to solve this problem an associative rule based approach is used to differentiate between the positive class and negative class member and to use this rules for segmentation.

Collecting customer data [9] and using the data for direct marketing operations have increasingly become possible. One approach is known as database marketing which is creating a bank of operation about individual customers from their orders, queries and other activities using it to analyze customer behavior and develop intelligent strategies [10],[13],[6].

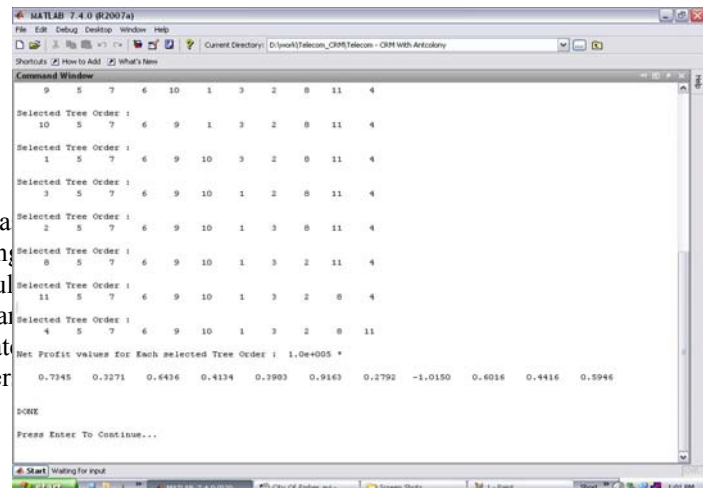
Another important computational aspect is to segment a customer group into sub-groups. AACO is a new technique that is used to find the accuracy in this paper.

All the above research works have aimed at finding a segmentation of the customer's database taking a predefined action for every customer based on that customer's current status. None of them have addressed about discovering actions that might be taken from a customer database, in this paper we have addressed about how to extract actions and find out the best accuracy for the customer to be loyal.

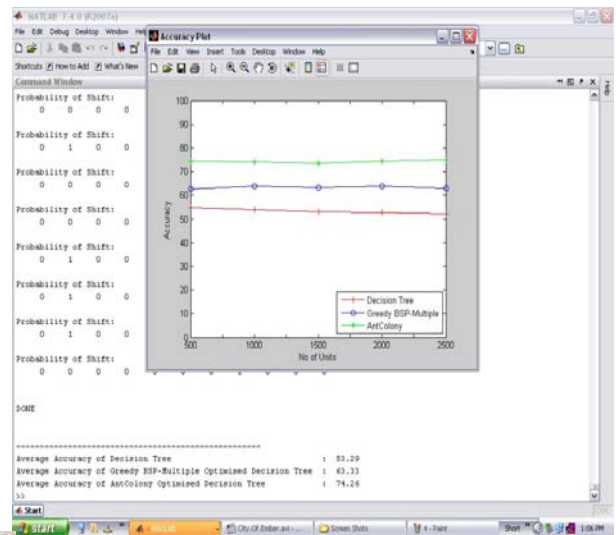
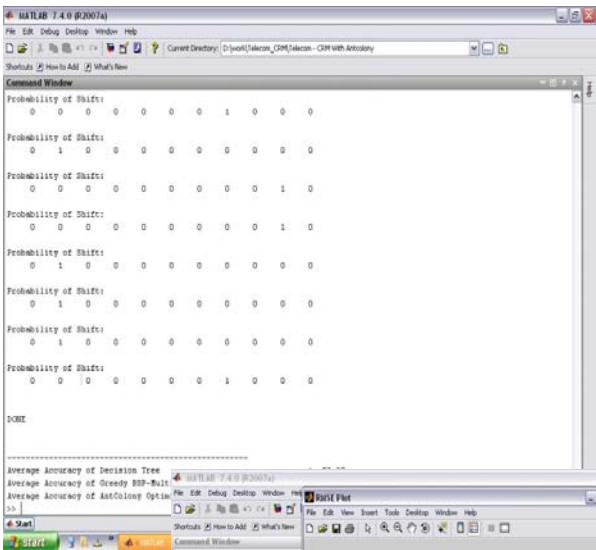
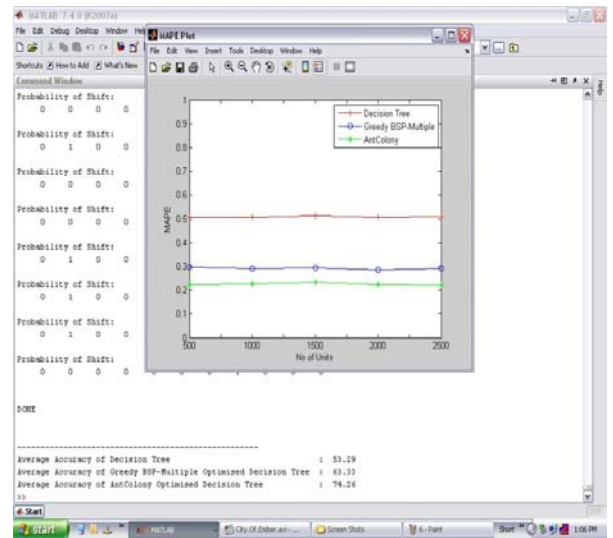
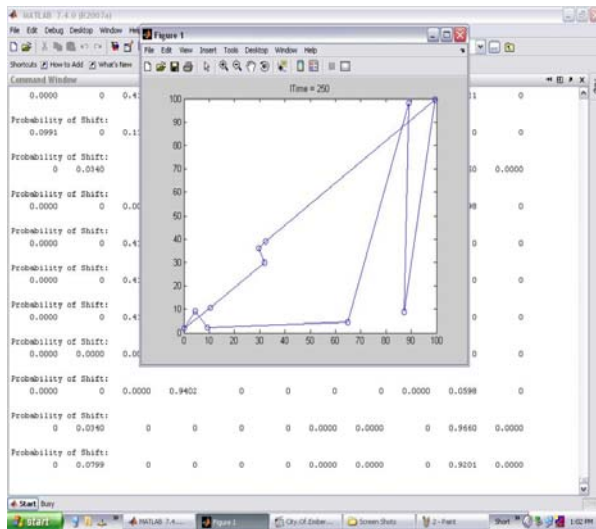
### EXPERIMENTAL EVALUATION



This experimental evaluation shows the entropy value that has been calculated for each parameter.



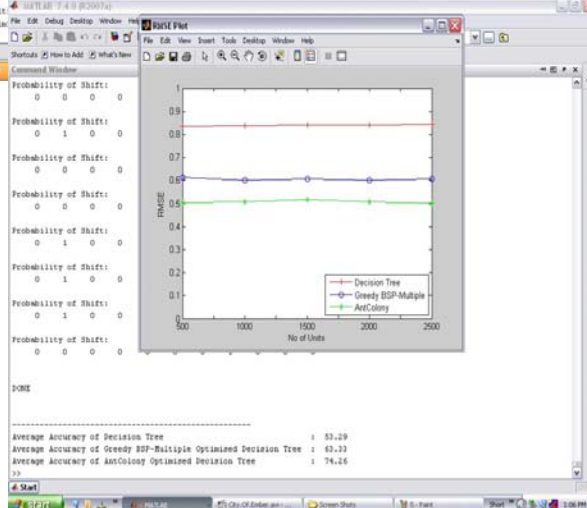
The best tree is selected from the various experimental result analyzed.



In our example each action set contains a number of actions and each action set contains four different actions on set with attribute changes. The experiment with Greedy BSP found action sets with maximum net benefit and is more efficient than optimal BSP. We also conducted this experiment with AACO and found that AACO is more accurate than the Greedy BSP algorithms.

## REFERENCES

- [1] R.Agarwal and R.Srikanth "Fast algorithms for the data bases (VLDB'94), pp.487-499, sept.1994.



- [2] Bank Marketing Association, Building a Financial Service Plan: Working plans for product and segment marketing, Financial Source Book, 1989.
- [3] A. Berson, K. Therling, and S.J. Smith, Building Data Mining Applications for CRM. McGraw-Hill, 1999.
- [4] Bezdek, J.C., (1981) Pattern Recognition with fuzzy Objective Function Algorithm”
- [5] M.S. Chen, J. Han and P.S. Yu, "Data Mining: An Overview from a Database Perspective," IEEE Trans on Knowledge and Data Engineering", 1996.
- [6] R.G. Drozdenko and P.D. Drake, "Optimal Database Marketing" 2002.
- [7] J. Hung and C.X. Ling, "Using Auc and Accuracy in Evaluating Learning Algorithms", IEEE Trans Knowledge and Data Engineering. Pp 299-310, 2005.
- [8] Lyche "A Guide to Customer Relationship Management" 2001.
- [9] H. Mannila, H. Toivonen, and I. Verkamo, "Efficient Algorithms for Discovering Association Rules" Proc workshop Knowledge discovery in databases.
- [10] E.L. Nash, "Database Marketing" McGraw Hill, 1993.
- [11] J.R. Quinlan, C4.5 Programs for Machine Learning", 1993
- [12] S. Rajasekaran, Vasantha Kalyani David , Pattern Recognition using Neural And Functional networks by 2008.
- [13] Rashes attrition and M. Stone DataBase Marketing, Joattritionhn Wiley, 1998.
- [14] Q. Yang, J. Yin, C.X. Ling, and T. Chen, "Post processing Decision Trees to Extract Actionable Knowledge" IEEE conf. Data Mining ,pp 685-688, 2003.
- [15] X. Zang and C.E. Brodley, "Boosting Lazy Decision Trees", Proc Conf .Machine Learning (ICM), pp 178-185, 2003.
- [16] Z.H. Zhou, J. Wu and W. Tang "Ensembling Neural Network IEEE Conf. Data Mining, pp 585-588, 2003.
- P. Senthil Vadivu received MSc (Computer Science) from Avinashilingam Deemed University in 1999, completed M.phil from Bharathiar university in 2006. Currently working as the Head, Department of Computer Applications, Hindusthan College of Arts and Science, Coimbatore-28. Her Research area is decision trees using neural networks.

# Constructing Models from Microarray Data with Swarm Algorithms.

Mrs.Aruchamy Rajini

Lecturer in Computer Applications  
Hindusthan College of Arts & Science, Coimbatore  
aruchamy\_rajini@yahoo.co.in

Dr. (Mrs.)Vasantha kalayani David

Associate Professor, Department of Computer Science  
Avinashilingam Deemed University, Coimbatore  
vasanthadavid@yahoo.com

## Abstract

Building a model plays an important role in DNA microarray data. An essential feature of DNA microarray data sets is that the number of input variables (genes) is far greater than the number of samples. As such, most classification schemes employ variable selection or feature selection methods to pre-process DNA microarray data. In this paper Flexible Neural Tree (FNT) model for gene expression profiles classification is done. Based on the pre-defined instruction/operator sets, a flexible neural tree model can be created and evolved. This framework allows input variables selection, over-layer connections and different activation functions for the various nodes involved. The FNT structure is developed using the Ant Colony Optimization (ACO) and the free parameters embedded in the neural tree are optimized by Particle Swarm Optimization (PSO) algorithm and its enhancement (EPSO). The purpose of this research is to find the model which is an appropriate model for feature selection and tree-based ensemble models that are capable of delivering high performance classification models for microarray data.

**Keywords** --- DNA, FNT, ACO, PSO, EPSO

## I. INTRODUCTION

A DNA micro array (also commonly known as DNA chip or gene array) is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip forming an array for the purpose of expression profiling, monitoring expression levels for thousands of genes simultaneously. Micro arrays provide a powerful basis to monitor the expression of thousands of genes, in order to identify mechanisms that govern the activation of genes in an organism [1].

Recent advances in DNA micro array technology allow scientists to measure expression levels of thousands of genes simultaneously in a biological organism. Since the cancer cells usually evolve from normal cells due to mutations in genomic DNA, comparison of the gene expression levels of cancerous and normal tissues or different cancerous tissues may be useful to identify those genes that might anticipate the clinical behavior of cancers.

Micro array technology has made the modern biological research by permitting the simultaneous study of genes comprising a large part of genome [2]. In response to the development of DNA micro array technologies, classification methods and gene selection techniques are

been computed for better use of classification algorithm in micro array gene expression data [3] [4].

Variable selection refers to the problem of selecting input variables that are most predictive for a given outcome. Appropriate variable selection can greatly enhance the effectiveness and potential interpretability of an inference model. Variable selection problems are found in all supervised and unsupervised machine learning tasks including classification, regression, time-series prediction, and clustering [5].

This paper develops a Flexible Neural Tree (FNT) [6] for selecting the input variables. Based on the pre-defined instruction/operator sets, a flexible neural tree model can be created and evolved. FNT allows input variables selection, over-layer connections and different activation functions for different nodes. The tuning of the parameters encoded in the structure is accomplished using Particle Swarm Optimization (PSO) algorithm and its enhancement.

The proposed method interleaves both optimizations. Starting with random structures and corresponding parameters, it first tries to improve the structure and then as soon as an improved structure is found, it then tunes its parameters. It then goes back to improving the structure again and, then tunes the structure and rules' parameters. This loop continues until a satisfactory solution is found or a time limit is reached.

## II. THE FLEXIBLE NEURAL TREE MODEL

The function set  $F$  and terminal instruction set  $T$  used for generating a FNT model are described as  $S = F \cup T = \{+_2, +_3, \dots, +_N\} \cup \{x_1, \dots, x_n\}$ , where  $+_i (i = 2, 3, \dots, N)$  denote non-leaf nodes' instructions and taking  $i$  arguments.  $x_1, x_2, \dots, x_n$  are leaf nodes instructions and taking no other arguments. The output of a non-leaf node is calculated as a flexible neuron model (see Fig.1). From this point of view, the instruction  $+_i$  is also called a flexible neuron operator with  $i$  inputs.

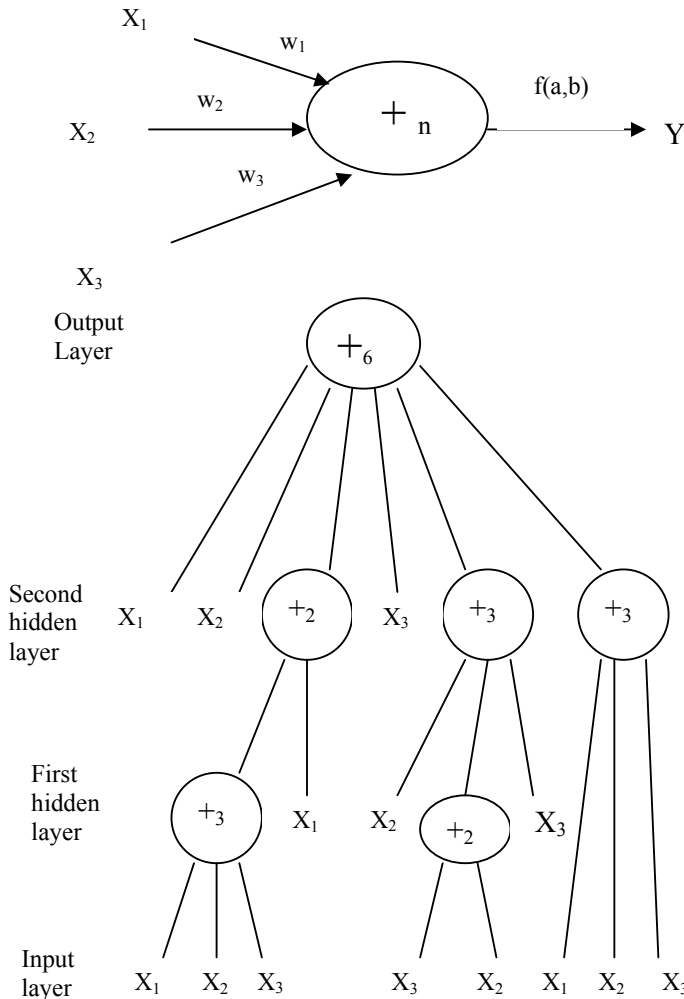
In the creation process of neural tree, if a nonterminal instruction, i.e.,  $+_i (i = 2, 3, 4, \dots, N)$  is selected,  $i$  real values are randomly generated and used for representing the connection strength between the node  $+_i$  and its children. In addition, two adjustable parameters  $a_i$  and  $b_i$  are randomly created as flexible activation function parameters and their value range are  $[0, 1]$ . For

developing the forecasting model, the flexible activation function  $f(ai, bi, x) = e^{-(x-ai)/bi)^2}$  is used.

The total excitation of  $+_n$  is

$$net_n = \sum_{j=1}^n w_j * x_j,$$

where  $x_j$  ( $j = 1, 2, \dots, n$ ) are the inputs to node  $+_n$  and  $w_j$  are generated randomly with their value range are  $[0,1]$ . The output of the node  $+_n$  is then calculated by  $out_n = f(a_n, b_n, net_n) = e^{-(net_n - a_n)/b_n)^2}$ . The overall output of flexible neural tree can be computed from left to right by depth-first method, recursively [7].



**Fig. 1.** A flexible neuron operator (left), and a typical representation of the FNT with function instruction set  $F = \{+_2, +_3, +_4, +_5, +_6\}$ , and terminal instruction set  $T = \{x_1, x_2, x_3\}$  (right)

### III. SWARM INTELLIGENCE ALGORITHMS.

Swarm Intelligence (SI) has recently emerged as a family of nature inspired algorithms, especially known for their ability to produce low cost, fast and reasonably accurate solutions to complex search problems [1]. It gives an introduction to swarm intelligence with special emphasis

on two specific SI algorithms well-known as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO).

PSO was originated from computer simulations of the coordinated motion in flocks of birds or schools of fish. As these animals wander through a three dimensional space, searching for food or evading predators, these algorithms make use of particles moving at velocity dynamically adjusted according to its historical behaviors and its companions in an n-dimensional space to search for solutions for an n-variable function optimization problem. The Particle Swarm Optimization algorithm includes some tuning parameters that greatly influence the algorithm performance, often stated as the exploration exploitation trade off. Exploration is the ability to test various regions in the problem space in order to locate a good optimum, hopefully the global one. Exploitation is the ability to concentrate the search around a promising candidate solution in order to locate the optimum precisely [8][9][10][11].

El-Desouky et al., in [10] proposed a more enhanced particle swarm algorithm depending on exponential weight variation instead of varying it linearly which gives better results when applied on some benchmarks functions. In this paper three models are compared: 1) A Tree structure is created with ACO 2) A Tree structure is created with ACO and the parameters are optimized with PSO 3) A Tree Structure is created with ACO and the parameters are optimized with EPSO. Comparisons of the three models are shown in this paper to propose an efficient methodology.

### IV. ANT COLONY OPTIMIZATION (ACO) FOR EVOLVING THE ARCHITECTURE OF FNT

ACO is a new probabilistic technique for solving computational problems to find optimal path. It is a paradigm for designing metaheuristic algorithm for combinatorial optimization problems. The main underlying idea, inspired by the behavior of real ants, is that of a parallel search over several constructive threads based on local problem data and on a dynamic memory structure containing information on the quality of previously obtained results.

In this algorithm, each ant will build and modify the trees according to the quantity of pheromone at each node. Each node memorizes the rate of pheromone. First, a population of programs is generated randomly. Each node is initialized at 0.5, which means that the probability of choosing each terminal and function is equal initially. The higher the rate of pheromone, the higher the probability to be chosen. Each ant is then evaluated using a predefined objective function which is given by Mean Square Error (MSE)[7].

$$Fit(i) = 1/p \sum_{j=1}^p (At - Ex)^2 \quad (1)$$



Where  $p$  is the total number of samples,  $At$  and  $Ex$  are actual and expected outputs of the  $j^{th}$  sample.  $Fit(i)$  denotes the fitness value of the  $i^{th}$  ant.

The pheromone is updated by two mechanisms:

– 1. Trail Evaporation: - Evaporation decreases the rate of pheromone for every instruction on every node, in order to avoid unlimited accumulation of trails, according to following formula:

$$Pg = (1 - \alpha) Pg - 1 \quad (2)$$

where  $Pg$  denotes the pheromone value at the generation  $g$ ,  $\alpha$  is a constant ( $\alpha = 0.15$ ).

– 2. Daemon actions: - For each tree, the components of the tree will be reinforced according to the Fitness of the tree. The formula is

$$P_{i,si} = P_{i,si} + \frac{\alpha}{Fit(s)} \quad (3)$$

where  $s$  is a solution (tree),  $Fit(s)$  its Fitness,  $si$  the function or the terminal set at node  $i$  in this individual,  $\alpha$  is a constant ( $\alpha = 0.1$ ),  $P_{i,si}$  is the value of the pheromone for the instruction  $si$  in the node  $i$ [7].

A brief description of AP algorithm is as follows:(1) every component of the pheromone tree is set to an average value; (2) random generation of tree based on the pheromone; (3) evaluation of ants (4) update of the pheromone; (5) go to step (1) unless some criteria is satisfied[7]

## V. PARAMETER OPTIMIZATION WITH PSO.

PSO [12] is in principle such a multi-agent parallel search technique. It does not require any gradient information of the function to be optimized, uses only primitive mathematical operators. Particles are conceptual entities which fly through the multi-dimensional search space.

PSO was inspired by the social behavior of a bird flock or fish school. PSO[13] conducts searches using a population of particles which correspond to individuals. In the PSO algorithm, the birds in a flock are symbolically represented as particles. These particles can be considered as simple agents flying through a problem space. A particle's location represents a potential solution for the problem in the multi-dimensional problem space. A different problem solution is generated, when a particle moves to a new location.

PSO model consists of a swarm of particles, which are initialized with a population of random positions. They move iteratively through the  $d$ -dimension problem space to search the new solutions, where the fitness,  $f$ , (Eqn. (1)) can be calculated as the certain qualities measure. Each particle has a position represented by a position-vector  $x_i$

( $i$  is the index of the particle) and a velocity represented by a velocity-vector  $v_i$ . Each particle remembers its own best position so far in a vector  $x$ .

Each particle keeps track of its own best position, which is associated with the best fitness it has achieved so far in a vector  $p_i$ . The best position among all the particles obtained so far in the population is kept track of as  $p_g$ .

Each particle  $i$  maintains the following information:  $x_i$  the current position of the particle,  $v_i$  the current velocity of the particle must be defined by parameters  $v_{min}$  and  $v_{max}$ . At each time step  $t$ , by using individual best position  $p_i$ , and all the global best position,  $p_g(t)$ , a new velocity for particle  $i$  is updated by[1]

$$V_i(t+1) = wv_i(t) + c_1\phi_1(p_i(t) - x_i(t)) + c_2\phi_2(p_g(t) - X_i(t)) \quad (4)$$

Where  $w$  is the inertia weight whose range is  $[0.4, 0.9]$ ,  $c_1$  and  $c_2$  are positive constant and are the learning factors called, respectively, cognitive parameter and social parameter. The proper fine-tuning may result in faster convergence and alleviation of local minima. The default values, usually,  $c_1=c_2=2$  are used. Even by using  $c_1=c_2=1.49$  gives better results.  $\phi_1$  and  $\phi_2$  are uniformly distributed random number in range of  $[0, 1]$ .

During the iteration time  $t$ , the update of the velocity from the previous velocity to the new velocity is determined. The new position is then determined by the sum of the previous and the new velocity, according to the formula:

$$X_i(t+1) = x_i(t) + V_i(t+1) \quad (5)$$

Various methods are used to identify particle to influence the individual. Two basic approaches to PSO exist based on the interpretation of the neighborhood of particles. They are (1) global best (gbest) version of PSO where the neighborhood of each particle is the entire swarm. The social component then causes particles to be drawn toward the best particle in the swarm.(2) local best (lbest) PSO model, particles have information only of their own and their nearest array neighbors best(lbest) rather than that of entire group. The gbest model converges quickly but has weakness of being trapped in local optima. The gbest is recommended strongly for unimodal objective function [1].

The PSO is executed with repeated application of equation (4), (5) until a specified number of iterations has been exceeded or when the velocity updates are close to zero over a number of iterations.

The PSO algorithm work as follows:

1) Initial population is generated randomly. The learning parameters  $c_1$ ,  $c_2$  are assigned in advance. 2) The objective function value for each particle is calculated. 3) Search point is modified. The current search point of each particle is changed using Equations (4) and (5). 4) If

maximum number of iterations is reached, then stop; otherwise go to step (2).

## VI. EXPONENTIAL PARTICLE SWARM OPTIMIZATION (EPSO)

In linear PSO, the particles tend to fly towards the gbest position found so far for all particles. This social cooperation helps them to discover fairly good solutions rapidly. However, it is exactly this instant social collaboration that makes particles stagnate on local optima and fails to converge at global optimum. Once a new gbest is found, it spreads over particles immediately and so all particles are attracted to this position in the subsequent iterations until another better solution is found. Therefore, the stagnation of PSO is caused by the overall speed diffusion of newly found gbest [10].

An improvement to original PSO is constituted by the fact that  $w$  is not kept constant during execution; rather, starting from maximal value, it is linearly decremented as the number of iterations increases down to a minimal value [4], initially set to 0.9, decreasing to 0.4 over the first 1500 iterations if the iterations are above 1500, and remaining 0.4 over the remainder of the run according to

$$W = (w - 0.4) \frac{(\text{MAXITER} - \text{ITERATION})}{\text{MAXITER}} + 0.4 \quad (6)$$

MAXITER is the maximum number of iterations, and ITERATION represents the number of iterations.

EPSO has a great impact on global and local exploration it is supposed to bring out the search behavior quickly and intelligently as it avoid the particles from stagnation of local optima by varying this inertia weight exponentially, as given

$$W = (w - 0.4) e^{(\text{MAXITER} - \text{ITERATION}) - 1} / \text{MAXITER} + 0.4 \quad (7)$$

By using the Equation (7) the movement of particles will be faster and distant from each other.

### A. General learning Procedure:

The general learning procedure for constructing the FNT model can be described as follows.

- 1) Create an initial population randomly (Set FNT trees and its corresponding parameters);
- 2) Structure optimization is achieved by the Ant Colony Optimization Algorithm.
- 3) If a better structure is found, then go to step 4), otherwise go to step 2);
- 4) Parameter optimization is achieved by the EPSO algorithm. In this stage, the architecture of FNT model is fixed, and it is the best tree developed during the end of run of the structure search. The parameters (weights and

flexible activation function parameters) encoded in the best tree formulate a particle.

5) If the maximum number of local search is reached, or no better parameter vector is found for a significantly long time then go to step 6); otherwise go to step 4);

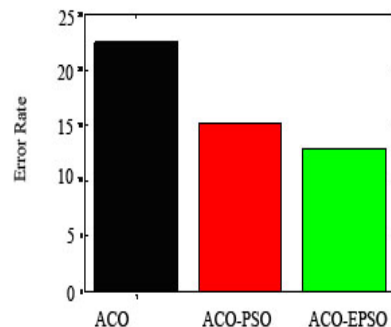
6) If satisfactory solution is found, its corresponding informative genes are extracted, then the algorithm is stopped; otherwise go to step 2).

## VII. RESULTS

As a Preliminary study, the Wisconsin Prognostic breast cancer (WPBC)[18] data set has 34 attributes (32 real-valued) and 198 instances. The methodology adopted for breast cancer data set was applied. Half of the observation was selected for training and the remaining samples for testing the performance of different models. All the models were trained and tested with same set of data. The instruction set used to create an optimal FNT classifier  $S = \text{FUT} = \{+2, \dots, +N\} \cup \{x_0, x_1, \dots, x_{31}\}$  Where  $x_i$  ( $i=0,1,\dots,31$ ) denotes the 32 input features. To get an optimal tree structure an ACO algorithm is applied. In this experiment the input is the number of ant and the number of iterations. Each ant is made to run for a specified number of iterations. Each ant constructs a neural tree with its objective function which is calculated as MSE. The ant which gives the low MSE is taken to be the best tree for which the parameters are optimized with PSO and EPSO. The tree which produces the low error is the optimized neural tree and this extracts the informative genes.

As with breast cancer data set, it was well proven that the tree structure with ACO and parameter optimization done with EPSO can achieve better accuracy compared with the other models. The main purpose is to compare the models quality, where the quality is measured according to the error rate, mean absolute percentage error and accuracy. The ACO-EPSO model has the smallest error rate when compared with the other models. All the three models are made to run for the same number of iterations and the results shows that ACO-EPSO success to reach optimal minimum in all runs. This method gives the best minimum points better than the other models. This is depicted in the following figures.

In Figure 1 and 2 the error rate and mean absolute percentage error of the model ACO-EPSO is low when compared with ACO and ACO-PSO.



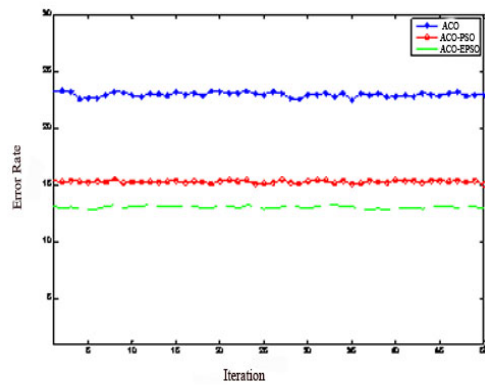


Fig1: Comparison of models in terms of error rate

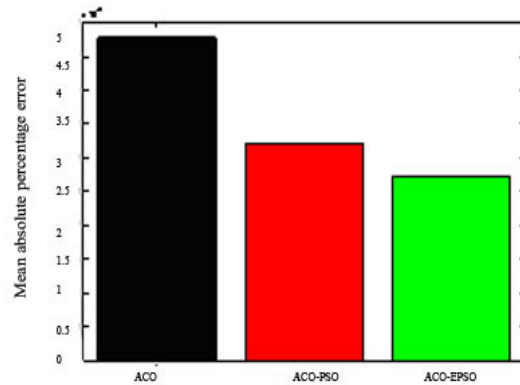


Fig2: Comparison of models in terms of mean absolute percentage error

In Figure 3 the accuracy of the model with ACO-EPSSO is high, which shows that the proposed model is highly efficient that it could be used for faster convergence and slower error rate.

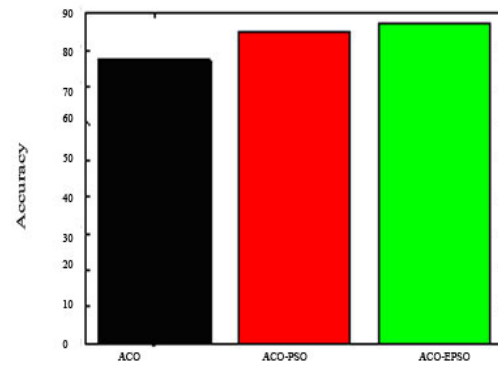
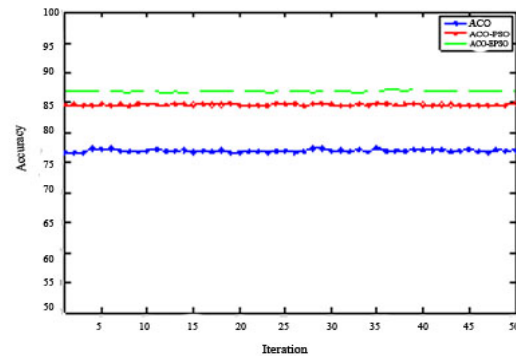


Fig3: Comparison of models in terms of accuracy



## VIII.CONCLUSION

A new forecasting model based on neural tree representation by ACO and its parameters optimization by EPSO was proposed in this paper. A combined approach of ACO and EPSO was encoded in the neural tree was developed. It should be noted that there are other tree-structure based evolutionary algorithms and parameter optimization algorithms that could be employed to accomplish same task but this proposed model yields feasibility and effectiveness. This proposed new model helps to find optimal solutions at a faster convergence. EPSO convergence is slower to low error, while other methods convergence faster to large error. The Proposed method increases the possibility to find the optimal solutions as it decreases with the error rate.

## REFERENCES

- [1] Swagatam DAjith Abraham ,Amit Konar, "Swarm Intelligence Algorithms in Bioinformatics", Studies in Computational Intelligence(SCI)94,113-147 Springer – Verlag Berlin Heidelberg (2008).
- [2] Per Broberg, "Statistical methods for ranking differentially expressed genes molecular sciences", AstraZeneca Research and Development Lund, s-221 87 Lund, Sweden 7 May 2003.
- [3] Hong Chai and Carlotta Domeniconi, "An Evaluation of Gene Selection methods for multiclass microarray Data Classification", Proceedings of the second European Workshop on Data Mining and Text Mining in bioinformatics.

- [4] J.Jager,R.Sengupta,W.L.Ruzzo,” *Improved gene selection for classifications of microarrays*”,Pacific Symposium on Biocomputing8:53-64(2003).
- [5] Yuehui Chen, Ajith Abraham and Bo Yang, “*Feature Selection & Classification using Flexible Neural Tree*”, Elsevier Science, 15 January 2006.
- [6] Chen, Y., Yang, B., Dong, J., “*Nonlinear systems modelling via optimal design of neural trees*”. International Journal of Neural Systems. 14, (2004) 125-138.
- [7] Yuehui Chen, Bo Yang and Jiwen Dong, “*Evolving Flexible Neural Networks using Ant Programming and PSO Algorithm*”, Springer – Verlag Berlin Heidelberg 2004.
- [8] Li-ping, Z., Huan-jun, Y., Shang-xu, H., *Optimal Choice of Parameters for Particle Swarm Optimization*, Journal of Zhejiang University Science, Vol. 6(A)6, pp.528-534, 2004.
- [9] Sousa, T., Silva, A., Neves, A., *Particle Swarm Based Data Mining Algorithms for Classification Tasks*, Parallel Computing 30, pp. 767-783, 2004.
- [10] El-Desouky N., Ghali N., Zaki M., *A New Approach to Weight Variation in Swarm Optimization*, proceedings of Al-azhar Engineering, the 9<sup>th</sup> International Conference, April 12 - 14, 2007.
- [11] Neveen I.Ghali, Nahed EL-Dessouki, Mervat A.N and Lamiaa Bakrawi, “*Exponential Particle Swarm Optimization approach for Improving Data Clustering*”, World Academy of Science, Engineering & Technology 42, 2008.
- [12] Kennedy.j, Eberhart.R. and Shi.Y. (2001),”*Swarm Intelligence*”,Morgan Kaufmann Academic Press.
- [13] Kennedy, J., Eberhart, R., *Particle Swarm Optimization*, Proceedings of the IEEE International joint conference on Neural networks, vol.4, pp. 1942-1948, 1995.
- [14] Shi, Y., Eberhart, R., *Parameter Selection in Particle Swarm Optimization*, proceedings of the 7th International Conference on Evolutionary Programming VII, pp. 591 – 600, 1998.
- [15]Vasanth Kalyani David, Sundaramoorthy Rajasekaran: “*Pattern Recognition using Neural and Functional Networks*” Springer 2009
- [16] Yuehui Chen, Ajith Abraham and Lizhi Peag, “*Gene Expression Profiling using Flexible Neural Trees*”, Springer – Verlag Berlin Heidelberg 2006.
- [17] Dr. V. Sarvanan, R. Mallika, “*An Effective Classification Model for Cancer Diagnosis using Micro Array Gene Expression Data*” IEEE xplore.
- [18] Available on the UW CS ftp server, ftp [ftp.cs.wisc.edu](ftp://ftp.cs.wisc.edu), cd math-prog/cpo-dataset/machine-learn/WPBC/
- [19] Yuehui CHEN,Ajith Abraham and Yong Zhang ,”Ensemble of Flexible Neural Trees for Breast Cancer Detection”2003.

## IMPROVED CONTENT BASED IMAGE RETRIEVAL USING COLOR HISTOGRAM AND SELF ORGANIZING MAPS

R.Priya  
HOD I/c  
Dept of Computer Science  
Sree Narayana Guru College  
Coimbatore, India.  
priyaminerva@rediffmail.com

Dr.VasanthaKalyaniDavid  
Associate professor  
Dept of Computer Science  
Avinashilingam Deemed University  
Coimbatore, India  
vasanthadavid@yahoo.com

### **Abstract**

Color is a feature of the great majority of content-based image retrieval systems. The conventional color histogram retrieval method is prone to lose the spatial information of colors. This paper proposes two methods; one combines color histograms with spatial information and the second which uses a dimensionality reduction technique that reduces the number of features. The experimental results show that the recall / precision and retrieval time of the proposed method is better than other methods.

**Keywords** – content-based image retrieval, color histogram, spatial information, Self Organizing Map

### **1. INTRODUCTION**

Content-Based Image Retrieval (CBIR) is defined as a process that searches and retrieves images from a large database on the basis of automatically-derived features such as color, texture and shape. The techniques, tools and algorithms that are used in CBIR, originate from many fields such as statistics, pattern recognition, signal processing, and computer vision. It is a field of research that is attracting professionals from different industries like crime prevention, medicine, architecture, fashion and publishing. The volume of digital images produced in these areas has increased dramatically over the past 10 decades and the World Wide Web plays a vital role in this upsurge. Several companies are maintaining large image databases, where the requirement is to have a technique that can search and retrieve images in a manner that is both time efficient and accurate (Xiaoling, 2009).

In order to meet these requirements, all the solutions, in general, perform the retrieval process in two steps. The first step is the 'feature extraction' step, which identifies unique signatures, termed as feature vector, for every image based on its pixel values. The

feature vector has the characteristics that describe the contents of an image. Visual features such as color, texture and shape are used more commonly used in this step. The classification step matches the features extracted from a query image with the features of the database images and groups' images according to their similarity.

Out of the two steps, the extraction of features is considered most critical because the particular features made available for discrimination directly influence the efficacy of the classification task (Choras, 2007). Out of the many feature extraction techniques, color is considered as the most dominant and distinguishing visual feature. A color histogram describes the global color distribution in an image and is more frequently used technique for content-based image retrieval (Wang and Qin, 2009) because of its efficiency and effectiveness. While the color histogram is robust to translation of object and rotation about the viewing axis, it does not include any spatial information. Moreover, due to its statistical nature, color histogram can only index the content of images in a limited way. This makes histogram inefficient while distinguishing images with same color but different color distributions. To avoid this, spatial information should be considered along with the color histogram (Rasheed *et al.*, 2008). This is termed as "spatial color histogram". Intuitively, the color histogram is a statistics on "how much" of each color contributing to the final histogram of an image, while a spatial distribution state of each color is a statistics on "where and how" the color is distributed in the image. Combining these two properties help to increase the accuracy of CBIR systems.

Several studies have been conducted to analyze the integration of spatial information with color histograms. Division of the whole histogram into region histograms using color clustering was considered by Pass and Zabih (1996). Hsu *et al.* (1995) modify the color histogram by first selecting a set of

representative colors and then analyzing the spatial information of the selected colors using maximum entropy quantization with event covering method. Colombo et al. (1998) propose a concept called Color Coherence Vector (CCV) to split histogram into two parts: coherent one and non-coherent one depending on the size of their connected component. Combining color with texture, shape and direction, this method escapes comparing color of different regions. Cinque et al. (1999) present a spatial-chromatic histogram considering the position and variances of color blocks in an image. Huang et al. (1997) proposes color correlogram for refining histogram which distills the spatial correlation of colors. Stricker and Dimai (1996) partition an image into 5 partially overlapping, fuzzy regions, extract the first three moments of the color distribution for each region, and then organize them into a feature vector of small dimension. Smith and Chang (1996) apply back-projection on binary color sets to extract color regions.

Recently, several other techniques have also been introduced where different color-related features are used as descriptors. Examples include the color feature hashing techniques [6], reference color table methods [7], and color adjacency graphs [8], etc. The sub-block histogram [9] retrieves images by separating an individual image into some blocks. All the above approaches improve the traditional histogram by embedding the spatial information into the color histogram. However, the way of extracting colors from the image spoils the robustness to rotation and translation of the conventional histogram as seen with the position of color block adopted by Cinque et al. [4], and the shape and size of the predefined triangle used by Rickman and John Stonham [10]. Therefore these improvement methods spoil the merit of the conventional histogram.

Xioling and Hongyan (2009) proposed a method to override the above difficulties by introducing a new method that combines color histogram and spatial information. The method was able to maintain the advantage of the robustness to image rotation and scaling of the traditional histogram, while incorporating the spatial information of pixels. They utilized groups of circular rings to segment the image first and then build color histograms over these rings into spatial space. This method had the advantage of having the perceptual sensitivity to the colors located in the central image by the use of a weighting factor of the histogram. This method has two issues. The first is the pixel size used for spatial integration and

another major difficulty associated with this method is its high dimensionality. Even with drastic quantization of the color space, the image histogram feature space occupied over 100 dimensions in real valued space. This high dimensionality indicates that methods of feature reduction can be implemented to improve the performance. Another side effect of large dimensionality is that it also increases the complexity and computation of the distance function. It particularly complicates 'cross' distance functions that include the perceptual distance between histogram bins [2]. In this paper, this method is enhanced to use a tree structured representation which combines color histogram and region features.

The paper is organized as below. Section 1 provided a brief introduction to the problem under discussion. Section 2 explains the base system as proposed by Xioling and Hongyan (2009). Section 3 explains the proposed methodology. Section 4 presents the experimental results while section 5 concludes the work.

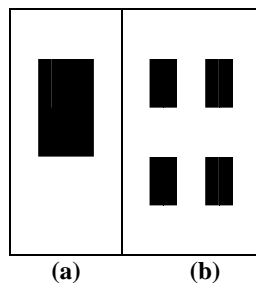
## 2. COMBINING HISTOGRAM AND SPATIAL INFORMATION

The system proposed by Xioling and hongyan (2009), referred to as base system in this paper, is explained in this section. Given a color space  $C$ , the conventional color histogram  $H$  of image  $I$  is defined as

$$H_c(I) = N(I, C_i) \quad (1)$$

where  $I$  is the image,  $C_i$  is the cell and  $I$  indicates the color levels in color space  $C$ ,  $N(I, C_i)$  denotes the number of pixels in  $i$  that fall into cell  $C_i$  and  $I \in [1..n]$ . From this definition, it can be seen that the conventional histogram is dependent on the summation of pixels of each color and ignores the spatial distribution of colors completely. Histograms only model the global color distribution of an image. Shuffling pixels within a given image does not change the histogram at all, even though the resultant image looks vastly different. For examples, consider Figures 1(a) and 1(b) where the total quantity of black color is equal. The color distribution is different but the color histogram considers only the total color amount and will fail to distinguish the two images.

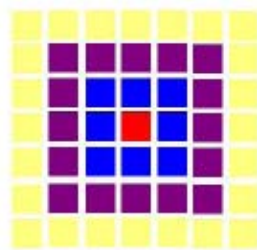




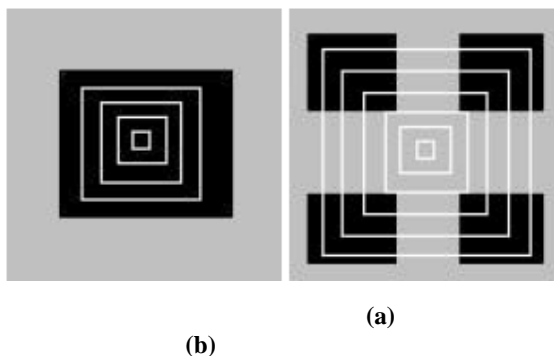
**Figure 1 : Example**

This was solved by dividing all the pixels contained in each bin of the color histogram in spatial space.

Groups of circular rings were used to segment the image and then for each region a color histogram in spatial space was built. An example of spatial division of the color histogram is shown in Figure 2. While using such a representation, Figure 1a and 1b will have their histograms as shown in Figure 3.



**Figure 2 : Spatial division of color histogram**



**Figure 3 : Histogram examples of Figures 1a and 1b**

The procedure starts by taking an image  $I$  of size  $M \times M$ , such that  $\text{Remainder} [(M-1)/2] = 0$  and  $M \geq 3$ . This condition makes sure that the location of the first ring starts at centre of the image  $I$ . Given a color space  $C$ , the histogram is defined as

$$H_C(I) = \{H_{R_i} \mid I \in [1..N]\} \quad (2)$$

where  $H_{R_i}$  is the color histogram of each ring. The number of rings is selected in such a way that it satisfies Equation 3.

$$\{ \max(n) \mid 0 < n \leq (M \times M)/8 + 1 \} \quad (3)$$

To measure two histograms, the frequently used Euclidean distance was used (Equation 4).

$$d(H(I), H(Q)) = \sum_{i=1}^n d(H_R(I), H_R(Q_i)) \quad (4)$$

To include perceptual sensitivity to the colors located in the centre of the image by using a weight factor, Equation 4 was modified to Equation 5.

$$d(H_R(I), H_R(Q)) = \left[ \sum_{j=1}^n W_j |H_R(I_j) - H_R(Q_j)|^2 \right]^{1/2} \quad (5)$$

where  $j$  indicates the color levels of color space  $C$ ,  $W_j$  is the  $j$ th weight of  $H_R$  and is calculated as

$$W_j = 1/j, \text{ for } j = 1 \dots n \quad (6)$$

and is normalized as

$$W_j = \frac{(1/j)}{\sum_{j=1}^n 1/j}$$

### 3. PROPOSED METHODOLOGY

The base system thus performed CBIR by considering the color distribution and layout and using a circular rings and calculating histograms for each ring. Color sensitivity is observed using a weighted scheme and Euclidean distance is used to measure the distance between two histograms. All the functions are performed on a HSV color space. Using this system, the authors were able to achieve reasonable accuracy over  $127 \times 127$  sized images. In the present work, the base system is improved in two manners.

**Method 1:** The base system used one pixel width for creating circular rings. When image size increases, the number of rings increases, thus increasing the number of histogram features. This increase has a direct impact on retrieval time which is more critical in online applications like the Internet and Satellite

based image retrieval systems. The proposed system attempts to reduce this feature space by increasing the width of the circular ring from one pixel to three pixels.

**Method 2:** The second method uses Self Organizing Map (SOM) to improve the accuracy, while simultaneously performs dimensionality reduction. The self-organizing Map (SOM), also known as a Kohonen map, is a technique which reduces the dimensions of data through the use of self-organizing neural networks. The proposed approach uses the improved histograms as image feature vector. The feature vector,  $F(I)$ , is generated for each image  $P$  in the collection. When a query is made, the feature vector for the query image is calculated and the similarity between any two images is calculated using the Euclidean distance between the feature vectors.

While combining the above procedure with SOM the following procedure is adhered. The SOM used consist of  $M \times M$  units, where  $M \times M \gg N$ . This makes it possible to map distinct feature vectors to unique locations in the SOM, by allowing each image to occupy its own region of the map. The weight vectors of all units in the SOM are stored using a single color histogram. Since each pixel can keep 3 values in its HSV color space to handle weight vectors of  $k$ -dimensions, the pixels are grouped in  $K/3$  tiles, with pixels from the same position of different tiles keeping the value of the same unit's weight vector. All values in the weight vectors are initialized using the weight factor calculated using Equation (5).

During training, the image blocks are given as input to the network. These input vectors are mapped with the network weight vectors to choose a neuron in the competitive layer as a winner. This winner is a neuron whose weight vector is much similar to the input vectors. In other words it is the neuron having the minimum Euclidean distance from the input vector. The input vector, say  $x$  is simultaneously applied to all nodes. The similarity between  $x$  and weight  $w_i$  is measured in terms of spatial neighborhood  $N_m$ . The weights affecting the currently winning neighborhood undergo adoption at the correct learning step other weights remain unaffected. The neighborhood,  $N_m$  is found around the best matching node  $m$  such that

$$\|x - w_m\| = \min [\|x - w_i\|]$$

The radius of  $N_m$  will be decreasing as the training progresses. Towards the end of training the neighborhood may involve no cells other than the central winning one. The

weight-updating rule for Self Organizing Feature Map is defined as

$$\Delta w_i(t) = \alpha[x(t) - w_i(t)] \text{ for } i \in N_m(t)$$

where  $N_m(t)$  denotes the current spatial neighborhood and  $\alpha$  denotes the learning rate. After training the weight vectors of each neuron of the Kohonen layer acts as code vectors.

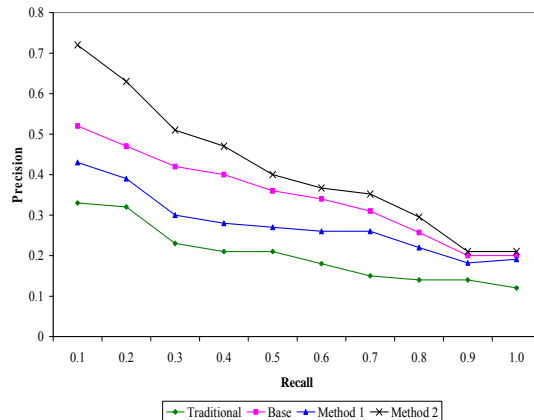
#### 4. EXPERIMENTAL RESULTS

The image database used during experimentation consists of 650 JPEG color images randomly selected from the World Wide Web. Figure 4 depicts a sample of images in the database. During testing, care was taken to choose a query image from different types of images like same scene, large change in appearance, etc. The performance metrics used during evaluation is the precision-recall measure and retrieval time. Precision is defined as the fraction of retrieved images that are truly relevant to the query image and recall is defined as the fraction of relevant images that are actually retrieved. Retrieval time is the time taken to retrieve images after giving the query image. The system was developed in MATLAB 7.3 and all the experiments were conducted in Pentium IV machine with 512 MB RAM. The histograms for all the images were constructed using 72 color bins after converting the RGB color space to HSV colour space.



**Figure 4: Sample Images from test database**

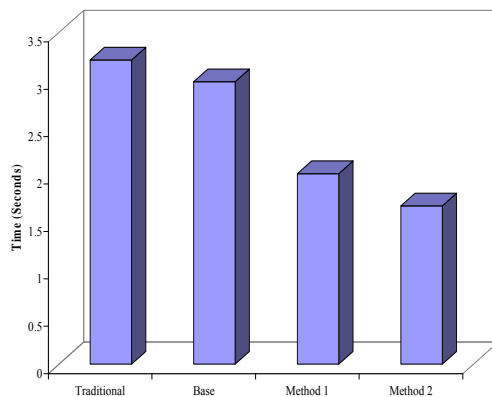
Figure 5 shows the precision-recall values obtained during testing for the two proposed systems. The results are compared with the base system as well as the traditional histogram method.



**Figure 5 : Precision and Recall**

From the figure, it could be seen that the proposed method which combines the base system with SOM outperforms the base, traditional histogram and Method 1 proposed systems in terms of precision and recall. The performance of Method 1 reduces the retrieval efficiency when compared with the base system. This result indicates that the combination of color distribution and spatial features with dimensionality reduction technique improves the retrieval of images, while just changing the pixel width does not.

The retrieval time taken by the four systems is shown in Figure 6.



**Figure 6 : Time Efficiency**

The retrieval time taken by both the proposed model is less than the traditional and base system. This shows that the dimensionality reduction in both cases is excellent. Method 1 when compared with traditional and base system showed 37.38 per cent and 32.55 per cent performance efficiency in terms of retrieval time. On the other hand showed 43.95 per cent and 16.92 per cent gain while comparing Method 2 with traditional and base systems respectively. The performance of Method 2 outperforms all the three systems.

## 5. CONCLUSION

This paper introduced methods to reduce the feature dimension space for content based image retrieval system while combining color space information with image spatial information. Two methods were proposed for this purpose. In the first method, the dimensionality reduction was performed by reducing the number of spatial histograms and the second method used SOM. From the results, it was found that the increase of circular ring width while successfully reduced the number of histogram features has a negative affect on the accuracy of the image retrieval system, with increase in speed of retrieval. The alternative solution using Self Organizing Map (SOM) for reducing the feature dimensionality proved to be the best in terms of precision/recall and retrieval time efficiency. Future research is planned to combine texture, shape features with color feature and SOM. Experiments to analyze the reason for the lower accuracy of Method 1 are also planned.

## References

- Choras, R.S. (2007) Image feature extraction techniques and their applications for CBIR and biometrics systems, International Journal of Biology and Biomedical Engineering, Vol.1, Issue 1, Pp. 6-16.
- Xiaoling, W. (2009) A Novel Circular Ring Histogram for Content-Based Image Retrieval, First International Workshop on Education Technology and Computer Science, Vol. 2, Pp.785-788.
- Wang, S. and Qin, H. (2009) A Study of Order-Based Block Color Feature Image Retrieval Compared with Cumulative Color Histogram Method, Sixth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD '09, IEEE Xplore, Vol. 1, Pp. 81 – 84.

Rasheed, W., An, Y., Pan, S., Jeong, I., Park, J. and Kang, J. (2008) Image Retrieval Using Maximum Frequency of Local Histogram Based Color Correlogram, Proceedings of the 2008 Second Asia International Conference on Modelling & Simulation (AMS), Pp. 322-326.

Pass, G. and Zabih, R. (1996) Histogram refinement for content-based image retrieval, Proc. WACV [C]. Sarasoto, FL, Pp. 96-102.

Hsu, wynne, Chua, T.S. and Pung, H.K.(1995) Integrated colorspatial approach to content-based image retrieval, Proceedings of the ACM International Multimedia Conference and Exhibition, Pp. 305-313.

Colombo, C., Del Bimbo, A. and Genovesi, I. (1998) Interactive image retrieval by color distributions, Proc. IEEE Int. Conf. on Multimedia Computing and Systems [C]. Texas, USA, Pp.255-258

Cinque, L., Levialdi, S., Olsen, K.A. and Pellicano, A. (1999) Color-based image retrieval using spatial-chromatic histograms, Proc. IEEE International Conference on Multimedia Computing and Systems [C]. Florence Italy, Pp. 969-973.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J. and Zabih, R. (1997) Image indexing using color correlograms, In Proc. IEEE Conf. Comp. Vision and Pattern Recognition [C], Pp.762-768.

Stricker, M. and Dimai, A.(1996) Color indexing with weak spatial constraints, SPIE proceedings, Vol. 2670, Pp.29 – 40.

Smith, J. and Chang, S.F. (1996) Tools and techniques for color image retrival. SPIE proceedings, Pp. 1630 –1639.

# Building an Energy-Efficient Prediction S-MAC Protocol for Wireless Networks

Mahmoud Abdel-Aziz El-Sakhawy Othman  
Information Technology Sector,  
Egyptian Financial Supervisory Authority (EFSA)  
Cairo, Egypt  
mahmoud.elsakhawy@efsa.gov.eg

Prof. Dr. Imane Aly Saroit Ismail  
Vice Dean for Education & Student Affairs,  
Faculty of Computers & Information, Cairo University  
Giza, Egypt  
iasi63@hotmail.com

**Abstract**—With the rapid development of wireless networking and micro-electro-mechanical systems (MEMS), wireless sensor networks (WSNs) have been immersed. WSNs consist of large amount of small, low-end, resource constrained devices, called sensors. Since sensor nodes are usually intended to be deployed in unattended or even hostile environments, it is almost impossible to recharge or replace their batteries. One of the most important research issues in the wireless sensor networks is to extend the network lifetime by energy efficient battery management. So, there are a lot of approaches that are designed to reduce the power consumption of the wireless sensor nodes. In this paper; a new protocol named "*prediction S-MAC protocol*" is proposed to reduce the power consumption of the wireless sensor nodes and to improve their performance compared to the previous S-MAC protocols.

**Keywords** - *Wireless sensor network; Sensor medium access control (S-MAC) protocol; periodic listen and sleep; adaptive listen, prolong listen, prediction S-MAC protocol.*

## I. INTRODUCTION

The term wireless networking refers to technology that enables two or more computers to communicate using standard network protocols and utilizing radio waves to maintain communication channels between computers.

Wireless sensor networking is an emerging technology that has a wide range of potential applications including environment monitoring, smart spaces, medical systems and robotic exploration. Such networks consist of large numbers of distributed nodes that organize themselves into a multi-hop wireless network [1 - 4].

Since wireless sensors are usually intended to be deployed in unattended or even hostile environments, it is almost impossible to recharge or replace their batteries [5]. The lifetime of a sensor node is much dependent on its power consumption. Hence, energy efficiency is of highly concern to the wireless sensor network design. So, there are a lot of approaches designed to reduce energy consumption in the wireless sensor networks [6 - 9]. Periodic listen and sleep protocol [6, 7], adaptive listen protocol [8] and prolong listen protocol [9] are examples of these protocols.

In the periodic listen and sleep protocol [6, 7], the time of each node is divided into two successive intervals; listen intervals and sleep intervals. In the sleep interval; a node sleeps

completely to preserve its power consumption. It turns off its radio and sets a timer to awake itself later to see if any other node wants to talk to it during listen time. This method decreases the average nodes power consumption, but increases average packets delay.

Adaptive listen protocol [8] is a modification of the periodic listen and sleep protocol; it reduces the packets' delay (*resulted in the periodic listen & sleep protocol*) by reducing the time spent in idle listen. Its basic idea is to let the node that is going to enter its sleep mode and overhears its neighbor's transmissions (ideally only RTS or CTS) wakes up for a short time at the end of the transmission. In this way, if the node is the destination node, its neighbour is able to immediately pass the data to it instead of waiting for its scheduled listen time.

Prolong listening protocol is proposed to improve the performance of the two previous protocols [9]. It also reduces the time spent in idle listen but by a greater value. It uses the concepts of both the periodic listen & sleep protocol and the adaptive listen protocol. In addition, if no RTS or CTS are heard before the node goes to its sleep mode, it sends a Ready To Receive (*RTR*) message to all its neighbours asking them if they are going to send in a short period of time (*prolong listen time*). If the node gets an answer, it exceeds its listen interval by a prolong listen time, on which it can send and receive instead of waiting for its scheduled listen time, so its neighbour is able to pass the data to it immediately instead of waiting for its scheduled listen time. If the node doesn't receive any answer, it will go to sleep until its next scheduled listen time. Results showed that prolong listen protocol increases both throughput and nodes life while decreases both delay and power consumption compared to periodic listen & sleep and adaptive listen protocols [9].

In this paper, a new S-MAC protocol named; "*prediction S-MAC protocol*" is proposed to improve the performance of the previous S-MAC protocols. Its basic idea is to divide the whole time of the node into two successive intervals; working interval (*listen interval*), in which the node is expected to send or receive packets and non-working interval (*sleep interval*), in which the node is not expected to send or receive packets.

The remainder of this paper is organized as follow: in the second section, medium access control for wireless sensor networks (*S-MAC*) and sources of energy waste in wireless

networks are illustrated. In addition, three existing S-MAC protocols; periodic listen & sleep protocol, adaptive listen protocol and prolong listen protocol are explained. The proposed prediction S-MAC protocol is explained and illustrated by detailed example in part three. Protocols implementations, parameters evaluation and results of the compared algorithms are discussed in part four. Finally, conclusion and future trends are given in section five.

## II. MEDIUM ACCESS CONTROL FOR WIRELESS SENSOR NETWORKS (S-MAC)

Medium Access Control (MAC) is a sub layer of the Data Link Layer of the seven layer Open Systems Interconnection (OSI) model. This layer is responsible for controlling the access of nodes to the medium to transmit or receive data. Sensor medium-access control protocols (S-MAC) are MACs designed for wireless sensor networks. The main task of the S-MAC protocol is to organize how the nodes in the WSN access the radio between nodes that are in radio range of each other. The most important attributes of S-MAC protocols to meet the challenges of the sensor network and its applications are; *collision avoidance*, *energy efficiency*, *scalability*, *channel utilization*, *latency*, *throughput and fairness* [6 - 8].

### A. Energy Efficiency in MAC Protocols

Energy efficiency is one of the most important issues in wireless sensor networks. To design an energy-efficient MAC protocol, the following question must be considered: what causes energy waste from the MAC perspective?. The following sources are major causes of energy waste [8]:

- **Collision** is the first source of energy waste. When two packets are transmitted at the same time and collide, they become corrupted and must be discarded. Follow-on retransmissions consume energy too. All S-MAC protocols try to avoid collisions one way or another.
- **Idle listening** happens when the radio is listening to the channel to receive possible data. The cost is especially high in many sensor network applications where there is no data to send during the period when nothing is sensed.
- **Overhearing** occurs when a node receives packets that are destined to other nodes. Overhearing unnecessary traffic can be a dominant factor of energy waste when traffic load is heavy and node density is high.
- **Control packet overhead** represents transmission and reception of control packets consume energy.
- **Overmitting** is the last source of energy waste, which is caused by the transmission of a message when the destination node is not ready. Given the facts above, a correctly-designed MAC protocol should prevent these energy wastes.

### B. Studied three existing S-MAC protocols

Sensor MAC protocols achieve an energy saving by controlling the radio to avoid or reduce energy waste from the above sources of energy waste. Turning off the radio when it is not needed is an important strategy for energy conservation. In this part, three existing S-MAC protocols are explained; periodic listen and sleep, adaptive listen and prolong listen. These

protocols have techniques in order to reduce the nodes' power consumption.

1) **Periodic Listen and Sleep Protocol:** This method was first proposed in [6] to reduce the power consumption of each node. It uses the fact that some nodes are idle for long time; means that the data rate is very low, so it is not necessary to keep nodes listening all the time. Periodic listen and sleep protocol reduces the listen time by putting nodes into periodic sleep state.

The basic scheme is shown in figure 1. Each node sleeps for some time, and then wakes up and listens to see if any other node wants to talk to it. During sleeping, the node turns off its radio, and sets a timer to awake it self later. A complete cycle of listen and sleep is called a *frame*. The listen interval is normally fixed according to physical-layer and MAC-layer parameters. The duty cycle is defined as the ratio of the listen interval to the frame length [7].

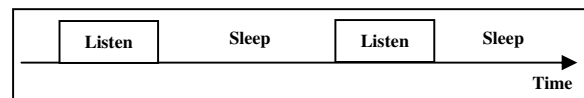


Figure 1. Periodic listen and sleep protocol.

All nodes are free to choose their own listen/sleep schedules, meaning that neighboring nodes may have different schedules. It should be noticed that not all neighboring nodes can synchronize together in a multi-hop network. Nodes exchange their schedules by periodically broadcasting a *SYNC* packet to their immediate neighbors, thus ensuring that all neighboring nodes can communicate even if they have different schedules. A node talks to its neighbors at their scheduled listen time, for example, if node A wants to talk to node B, it must wait until B is listening.

**Advantage:** The scheme of periodic listen and sleep is able to significantly reduce the time spent on idle listening when traffic load is light, so the power consumption is reduced.

**Disadvantage:** The downside of the scheme is the increased delay due to the periodic sleeping, which can accumulate on each hop.

2) **Adaptive listen Protocol:** The adaptive listen protocol was proposed in [7] to improve the delay caused by the periodic sleep of each node in a multi-hop network. It is modification of the periodic listen and sleep protocol, the basic idea is to let the node whose sleep interval is about to start and overhears its neighbor's transmissions (*ideally only RTS or CTS*) wakes up for a short period of time at the end of the transmission. In this way, if the node is the destination node, its neighbor will be able to immediately pass the data to it instead of waiting for its scheduled listen time, other nodes will go back to sleep until its next scheduled listen time. SYNC packets are sent at scheduled listen time to ensure all neighbors can receive it.

For example in figure 2, nodes 2 and 7 are about to enter their sleep interval, but node 1 has a packet to send to node 2, so it sends a *RTS*. All nodes in node 1's range; 2, 7 and 9 hear the



transmission so nodes 2 and 7 will extend their listen interval to receive the RTS (node 9 is already in the listen interval). After receiving the RTS, node 2 will extend its listen interval to serve the packet (sends CTS, receives data and sends an ACK), while node 7 doesn't have to extend its listen interval any more, so it enters its sleep interval.

**Advantage:**

Adaptive listen protocol improves throughput and decreases delay & power consumption compared to periodic listen and sleep protocol.

**Disadvantage:**

Since any packet transmitted by a node is received by all its neighbours even though only one of them is the intended receiver, it is clear that all nodes that overhear their neighbour's transmissions (RTS or CTS) wake up until they discover that the transmission is not for them although only one node is intended.

3) *Prolong Listen Protocol*: Prolong listening protocol is proposed in [9], which is a modification of both the periodic listen and sleep and adaptive listening protocols to improve their performance. This method takes the benefits of the two previous methods: first, it uses periodic listen and sleep concept and second, nodes that overhear RTS or CTS from its neighbors extend its listen interval to be able to receive packets instead of letting them wait for its scheduled listen time. The new part is; if no RTS and CTS are heard before the node goes to its sleep mode, it sends a ready to receive (RTR) message to all its neighbors asking them if they are going to send in a short period of time (prolong listen time). If the node gets an answer, it will exceeds its listening interval by a prolong listen time, on which it can send and receive, so its neighbor is able to immediately pass the data to it instead of waiting for its scheduled listen time. If the node doesn't receive any answer, it will go to sleep until its next scheduled listen time [9].

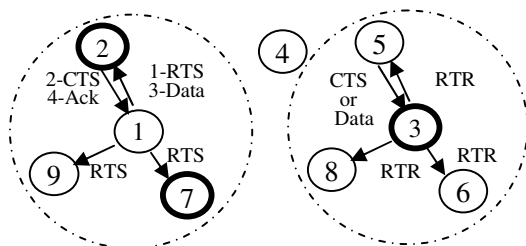


Figure 3. Prolong listen protocol.

For example, in figure 3, nodes 2, 3 and 7 are about to enter their sleep interval. But since node 2 hears a RTS from node 1, so it extends its listen interval to serve the packet as in the adaptive listen protocol. While node 3 hears nothing, so it sends a RTR message to all its neighbors. All nodes in node 3's range; 5, 6 and 8 hear the transmission. Node 5 responds (by sending a RTR reply or by just sending the data), so node 3 prolong its

listen interval to serve the packet. Note that, node 4 does nothing because it is out of range.

**Advantage:** Since prolong listen protocol services a lot of packets during prolong listen time instead of letting them wait for their next scheduled listen time, so it improves throughput and decreases delay and power consumption compared to periodic listen and sleep protocol and adaptive listening protocol.

**Disadvantage:** It is clear that all nodes that overhear their neighbour's transmissions (RTS or CTS) wake up until they discover that the transmission is not for them although only one node is intended.

It should be noted that not all next-hop nodes can overhear a RTR message from the transmitting node because they are not at the scheduled listen time or they do not have data packets to send. So if a node starts a transmission by sending out an RTR message during prolong listen time, it might not get a reply. In this case, it just goes back to sleep and will try again at the next normal listen time and a RTR message consume energy too.

### III. PROPOSED PREDICTION S-MAC PROTOCOL

The previous S-MAC protocols are based on initial listen & sleep intervals, where listen time in each frame is fixed usually about 300 msec, while the sleep time can be changed to reflect different duty cycles. The downside of the previous schemes is the increased delay due to the periodic sleeping which is accumulated on each hop. In addition, during listen intervals, nodes may have no data to transmit / receive (idle) or service their data in a partial time of the listen intervals. These techniques imply to minimize the sensor node lifetime.

In this section; we propose a new protocol named "*prediction S-MAC protocol*" to handle the problems of the previous S-MAC protocols. It does not depend on fixed listen and sleep intervals. Instead the node transmits only (send/receive) according to the prediction of its listen intervals, otherwise it goes to sleep mode and turns off its radio until expectation of its next listen interval. The basic idea of the proposed protocol is to divide the whole time of the node into two successive intervals; working interval (listen interval), in which the node is expected to send or receive packets and non-working interval (sleep interval), in which the node is not expected to send or receive packets.

Confidence interval method is used to predict the working and non-working intervals based on the last previous  $N$  listen (working) intervals. It is expected that the proposed prediction S-MAC protocol will increase both throughput and nodes' life, while it will decrease both delay and power consumption compared to the prolong listen protocol which was considered as the best protocol of the existing S-MAC protocols.

#### A. Parts of the proposed protocol

Proposed prediction S-MAC protocol consists of the following parts; non-sleep periods, prediction S-MAC intervals, packets arrival and adaptive listen / sleep as shown in figure 4. Prediction S-MAC intervals part consists of two steps; confidence interval calculation and expected listen & sleep intervals. In the following steps, parts of the proposed prediction S-MAC protocol are explained.

- **Non-sleep periods**

In the non-sleep periods, nodes are always in active mode (transmit, receive or idle state) without sleep time, take the first ( $N$ ) listen (send / receive) intervals in order to predict the next listen intervals.

- **Prediction S-MAC intervals**

In this part, listen & sleep intervals are expected based on the last previous ( $N$ ) listen intervals. Confidence interval method is used to predict the listen intervals by calculating both the mean and variance of the last previous ( $N$ ) listen intervals. Then law of large numbers of the ratios 90 %, 95 %, 99 % (or whatever) is used

to predict the start and end calculated confidence listen intervals. Then lower and upper bounds of the expected listen intervals are determined by adding both the start & end calculated confidence listen intervals to the upper bounds of the last previous intervals. Sleep intervals can be also expected. This part is divided into two steps:

1) **Confidence interval calculation:**

In this step, confidence interval (C.I) method is used to expect the listen intervals based on the last previous ( $N$ ) listen intervals as following:

- As known, confidence interval method gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data [10]. So, by using the confidence interval method the next listen interval ( $N+1$ ) can be expected based on the last previous ( $N$ ) listen intervals by the chosen ratios of 90%, 95%, 99% (or whatever) using the mean and variance of these ( $N$ ) listen intervals.

- To expect the listen interval ( $N+1$ ) based on the last previous ( $N$ ) listen intervals (Non-sleep periods) do the following;

- Compute both the mean and variance of these ( $N$ ) listen intervals where;

$$\bar{L}_{N_0} = \frac{\sum_{j=1}^N \text{listen}_j}{N} \quad (1)$$

where,  $N$ : is the previous listen intervals used,  
 $\text{listen}_j$ : is the listen period of interval  $j$ .

- Variance ( $S_0^2$ ) of these ( $N$ ) listen intervals is;

$$S_0^2 = \frac{Y_{N_0}}{N} - (\bar{L}_{N_0})^2 \quad (2)$$

$$\text{where, } Y_{N_0} = \sum_{j=1}^N (\text{Listen}_j)^2 \quad (3)$$

- By using law of large numbers and substituting into the following equation to compute both the start & end calculated confidence listen interval ( $N+1$ ) [10];

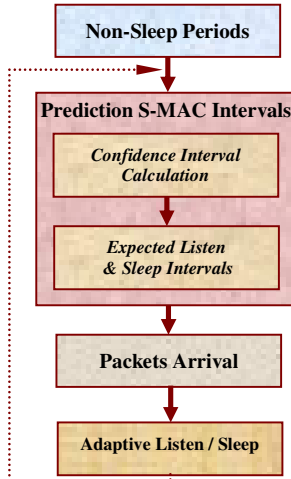


Figure 4. Parts of the prediction S-MAC algorithm

- Start and end calculated confidence listen interval ( $N+1$ )

$$\text{are } \bar{L}_{N_0} \mp m * \frac{S_0}{\sqrt{N}} \quad (4)$$

where,  $m = 1.65$  (using C.I of 90 %)  
 $= 1.96$  (using C.I of 95 %)  
 $= 2.58$  (using C.I of 99 %),

$S$  is the standard deviations  $= \sqrt{S^2}$ .

- Similarly to expect the listen intervals ( $i$ ) based on the last previous ( $N$ ) listen intervals, where  $i = N+2, N+3, \dots$ , and so on.

- Update both the last value of mean & variance by adding the last previous expected listen interval ( $i-1$ ) and excluding the previous listen interval ( $i-N-1$ ) where;

$$\bar{L}_{N_i} = \bar{L}_{N_{(i-1)}} - \frac{\text{listen}_{(i-N-1)}}{N} + \frac{\text{listen}_{(i-1)}}{N} \quad (5)$$

$$S_i^2 = \frac{Y_{N_i}}{N} - (\bar{L}_{N_i})^2 \quad \text{Where,} \quad (6)$$

$$Y_{N_i} = Y_{N_{(i-1)}} - (\text{Listen}_{i-N-1})^2 + (\text{Listen}_{i-1})^2 \quad (7)$$

- Calculate both the start and end calculated confidence interval ( $i$ ) where;

- Start & end calculated confidence listen interval ( $i$ )

$$\text{are } \bar{L}_{N_i} \mp m * \frac{S_i}{\sqrt{N}} \quad (8)$$

2) **Expected listen & sleep intervals:**

In this step, both the lower & upper pounds of the expected listen intervals are computed as following:

- After determining both the start & end calculated confidence listen interval ( $i$ ), the lower & upper bounds of the expected listen interval ( $i$ ) are expected by adding both the start & end calculated confidence listen interval ( $i$ ) to the upper bound of the last previous interval ( $i-1$ ).

- Lower & upper bounds of the sleep intervals can be also expected.

- **Packets arrival**

- a. If the arrival packets are in the expected listen interval (expected by 95 % or 99 %);

- Send the packets.
- Extend listen time, if transmission time is more than the expected listen interval.

- b. If the arrival packets are in the expected sleep interval (expected by 5% or 1 %);

- Do not send the packets.
- Reschedule the packets start time to the next predicted listen time.

- **Adaptive listen / sleep**

Since transmit time, receive time, idle time and sleep time of each node in the prediction S-MAC protocol are needed to evaluate the proposed protocol. Therefore, these times are assigned by adaptation listen (send / receive) intervals of the non-sleep periods according to the expected listen & sleep intervals of the prediction S-MAC intervals.

### B. Example of the proposed protocol

In the following example, steps of working the proposed prediction S-MAC protocol are illustrated.

1) *Non-sleep Periods*: In the *non-sleep periods*, nodes are always in active mode (*transmit, receive or idle* state) without sleep time. Suppose a network includes node (N1) which has the following transmissions with the other nodes (N2, N3 and N4). Figure 5.a shows a sample from node (N1) transmissions while figure 5.b illustrates send, receive and idle periods of node (N1).

Source	Destination	Status	Start time	Packets length
1	2	send	5	15
3	1	recv	25	10
1	4	send	45	13
1	3	send	65	25
2	1	recv	95	10
4	1	recv	115	5
1	3	send	127	8
3	1	recv	140	10
1	2	send	163	12

Figure 5.a. A sample from transmissions concerning node N1.

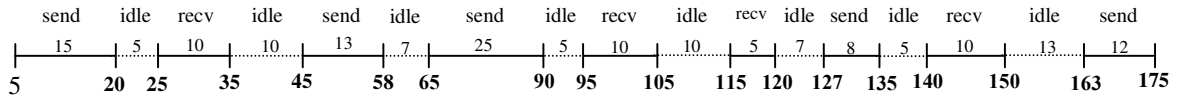


Figure 5.b. Send, receive and idle periods of node N1.

Figure 5. Non-sleep periods.

2) *Prediction S-MAC Intervals*: In this part, listen and sleep intervals are expected based on the last previous ( $N$ ) listen intervals. It is divided into two steps; *confidence interval (C.I) calculation* and *expected listen & sleep intervals*.

a) *Confidence interval (C.I) calculation*: Confidence interval method is used to calculate both the *start & end* calculated confidence listen intervals by calculating both the *mean & variance* of the last previous ( $N$ ) listen intervals. Then, law of large numbers of the ratio 95% is used in order to calculate both the *start & end* calculated confidence listen intervals. For example, fifth listen interval can be calculated based on the first four listen (*send, receive*) intervals by calculating both the *mean & variance* of these first four *send/receive* intervals of the *non-sleep periods*. Then, substituting in law of large numbers of the ratio 95% to get the *start & end* calculated confidence fifth listen interval as follows;

- The listen intervals of the first four listen (*send, receive*) intervals of the *non-sleep periods* are;  $L_0 = 15, 10, 13, 25$  ms.

- both the mean & variance of these listen intervals are;

$$\bar{L}_0 = \frac{(15 + 10 + 13 + 25)}{4} = 15.75, \quad (\bar{L}_0)^2 = 248.1$$

$$- Y_0 = (15)^2 + (10)^2 + (13)^2 + (25)^2 \\ = 225 + 100 + 169 + 625 = 1119,$$

$$- S_0^2 = \frac{Y_0}{N} - (\bar{L}_0)^2 = \frac{1119}{4} - 248.1 = 31.65, \quad S_0 = 5.6$$

- Start calculated confidence listen interval =

$$\bar{L}_0 - m * \frac{S_0}{\sqrt{N}} = 15.75 - 1.96 * \frac{5.6}{2} \approx 10$$

- End calculated confidence listen interval =

$$\bar{L}_0 + m * \frac{S_0}{\sqrt{N}} = 15.75 + 1.96 * \frac{5.6}{2} \approx 21$$

By updating both the *mean & variance* of the last expected listen interval (*fifth listen interval*), the *start & end* calculated confidence of six listen interval can be calculated based on the last previous four listen intervals as following;

- The last four previous listen intervals are;

$L_6 = 10, 13, 25, 11$  msec. (where 11 is the fifth listen interval calculated from the first four send/receive intervals of the non-sleep periods).

- Calculate both the *mean & variance* of the last four previous listen intervals where;

$$- \bar{L}_6 = 15.75 - \frac{15}{4} + \frac{11}{4} = 14.75, \quad (\bar{L}_6)^2 = 217.6$$

$$- Y_6 = 1119 - (15)^2 + (11)^2 = 1119 - 225 + 121 = 1015$$

$$- S_6^2 = \frac{Y_6}{N} - (\bar{L}_6)^2 = \frac{1015}{4} - 217.6 = 36.15,$$

$$S_6 = \sqrt{36.15} = 6.01$$

- Start calculated confidence listen interval =  $\bar{L}_6 - m * \frac{S_6}{\sqrt{N}}$

$$= 14.75 - 1.96 * \frac{6.01}{2} \approx 9$$

- End calculated confidence listen interval =  $\bar{L}_6 + m * \frac{S_6}{\sqrt{N}}$

$$= 14.75 + 1.96 * \frac{6.01}{2} \approx 21$$

b) *Expected listen & sleep intervals*: In the expected listen and sleep intervals, *lower & upper* bounds of the expected listen intervals are obtained by adding the *start & end* calculated confidence listen intervals to the upper bounds of the last previous intervals. In the previous example, after expecting both the *start & end* calculated confidence listen intervals, both the *lower & upper* bounds of the listen & sleep intervals are predicted as following;

- *Lower bound* of the expected fifth listen interval =  $90 + 10 = 100$  msec. (where 90 is the upper bound of the fourth listen interval as appearing in the figure 5).
- *Upper bound* of the expected fifth listen interval =  $90 + 21 = 111$  msec.
- Therefore, the *lower & upper* bounds of the expected fifth listen interval are 100 msec and 111 ms respectively,  $\Delta t = 11$  msec.
- Also, *lower & upper* bounds of the expected sleep interval are 90 msec and 100 msec respectively as shown in figure 6.

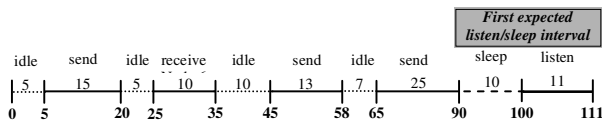


Figure 6. First expected Listen /sleep interval

- Also, *lower bound* of the expected six listen interval =  $111 + 9 = 120$  msec. (where 111 is the upper bound of the fifth listen interval expected as shown in figure 6).
- *Upper bound* of the expected six listen interval =  $111 + 21 = 132$  msec.
- Therefore, *lower & upper* bounds of the expected six listen interval are 120 msec and 132 msec respectively,  $\Delta t = 12$  msec.
- Also, *lower & upper* bounds of the expected sleep interval are 111 msec and 120 msec respectively as shown in figure 7.

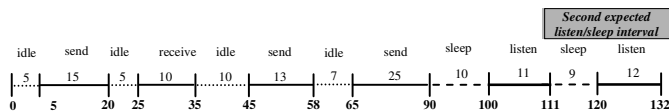


Figure 7. Second expected listen /sleep interval

By applying the previous steps, the seventh and eighth listen / sleep intervals can be expected where;

- Third expected listen/sleep interval (*seventh listen /sleep interval*) is:
  - *Start & end* calculated confidence listen interval = (10, 21).
  - *Lower bound* of the expected seventh listen interval =  $132 + 10 = 142$  msec.
  - *Upper bound* of the expected seventh listen interval =  $132 + 21 = 153$  msec.
  - Therefore, *lower & upper* bounds of the expected seventh listen interval are 142 ms and 153 ms respectively,  $\Delta t = 11$  ms.
  - Also, *lower & upper* bounds of the expected sleep interval are 132 ms and 142 ms respectively as shown in figure 7.
- Fourth expected listen/sleep interval (*eighth listen/sleep interval*) is:
  - *Start & end* calculated confidence listen interval = (9, 21)
  - *Lower bound* of the expected eighth listen interval =  $153 + 9 = 162$  msec.
  - *Upper bound* of the expected eighth listen interval

=  $153 + 21 = 174$  msec.

- Therefore, *lower & upper* bounds of the expected eighth listen interval are 162 msec and 174 msec respectively,  $\Delta t = 12$  msec.
- Also, *lower & upper* bounds of the expected sleep interval are 153 msec and 162 msec respectively as shown in figure 8.
- Therefore, listen/sleep intervals of the prediction S-MAC intervals are shown in figure 8.

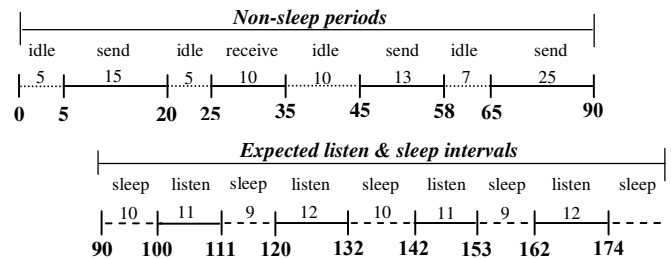


Figure 8. Expected listen & sleep intervals

Note that, nodes in the *prediction S-MAC intervals* are almost active during the expected listen time otherwise; nodes are almost in a sleep mode to serve the radio power consumption. Also, the arrival packets are almost in the expected listen interval by the ratio 95% to be served. We expect the *proposed prediction S-MAC protocol* will improve the performance parameters evaluation; average packet delay, throughput, average node power consumption and sensor node life compared to the *prolong listen protocol*.

### 3) Packets arrival:

- If the arrival packets are in expected listen interval;
  - Send the packets.
  - Extend listen time, if transmission time is more than the expected listen interval as shown in figure 9.
- If the arrival packets are in expected sleep interval;
  - Do not send the packets.
  - Reschedule the packets start time to the next expected listen time as shown in figure 9.

### 4) Adaptive listen / sleep periods:

To measure the performance parameters of the prediction S-MAC protocol; *average packet delay*, *throughput*, *average node power consumption* and *average node life*, we need to calculate both *transmit time*, *receive time*, *idle time* and *sleep time* for each node in the proposed protocol. These times are assigned by adaptation the listen (*send / receive*) intervals of the non-sleep periods according to the expected listen & sleep intervals of the prediction S-MAC intervals. Therefore, *transmit time*, *receive time*, *idle time* and *sleep time* of each node in the proposed prediction S-MAC protocol are assigned accurately as shown in figure 9.

As shown in figure 9, transmit time, receive time, sleep time and idle time of the node (N1) in the used example can be measured as follows;

- *Transmitting time* =  $15 + 13 + 25 + 8 + 12 = 73$  msec.
- *Receiving time* =  $10 + 11 + 5 + 10 = 36$  msec.
- *Sleep time* =  $10 + 9 + 7 + 9 = 35$  msec.
- *Idle time* =  $5 + 5 + 10 + 7 + 1 + 2 + 1 + 1 = 32$  msec.

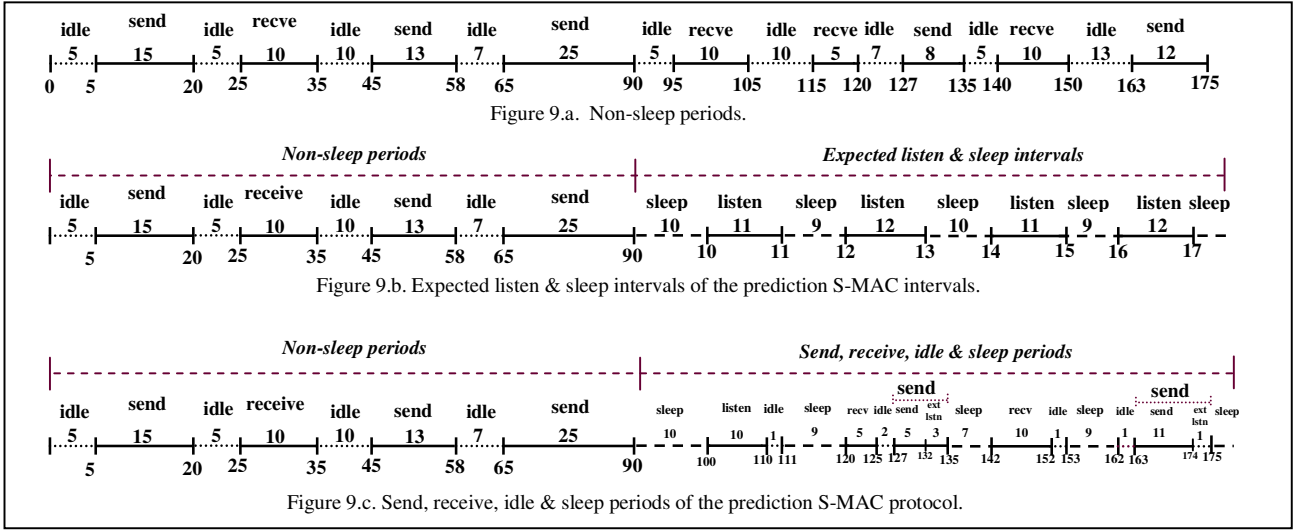


Figure 9. Adaptive listen / sleep concerning node (N1).

#### IV. PROTOCOLS IMPLEMENTATION

A simulation program is built using "Visual C++ programming language" in order to build and compare both the proposed prediction S-MAC protocol and the prolong listen protocol. The simulation program is divided into seven main parts; *packets creation*, *non-sleep periods*, *prediction S-MAC intervals*, *packets arrival*, *adaptive listen/sleep*, *prolong listen protocol* and *performance parameters evaluation* for each protocol. Prediction S-MAC intervals part consists of two steps; confidence interval calculation and expected listen & sleep intervals as shown in figure 10.

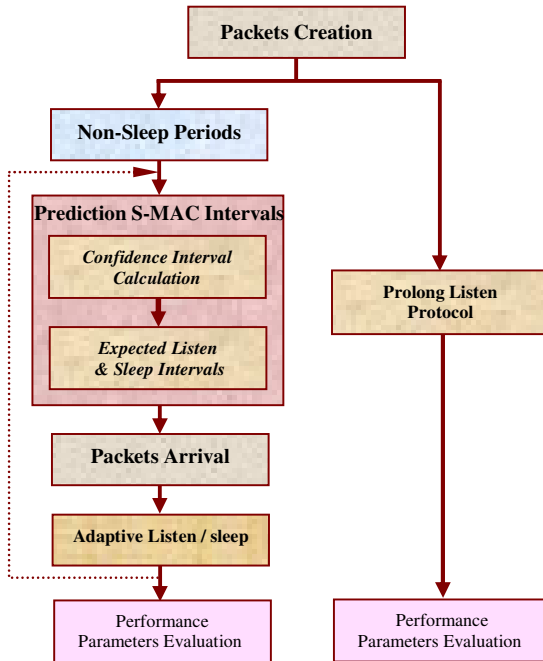


Figure10. Main parts of the simulation program

In order to obtain accurate results (*similar to real cases*), the packets' information is created randomly. However, the same packets' information must be used for the compared protocols in order to obtain accurate results. Therefore, the packets' creation part is separated from the parts of the compared S-MAC protocols, in fact its outputs is considered as common inputs for the compared S-MAC protocols.

Packets creation intervals are considered a common part for both the compared algorithms with generated values and the inputs for the S-MAC protocols. Of course; using the same input values for the compared protocols allows real comparison between them.

##### A. Performance parameters evaluation

The following parameters are used to evaluate both the *proposed prediction S-MAC* protocol and the *prolong listen* protocol; average packet delay, throughput, average node power consumption and average node life. A proposed protocol's objective is to increase both throughput and average node power consumption while decreasing both delay and average node power consumption.

##### 1) Average packet's delay:

Packet delay refers to the delay from when a sender has a packet to send until the packet is successfully received by the receiver. In sensor networks, the importance of delay depends on the application. Of course, the previous S-MAC protocols have longer delay due to the periodic sleeping on each hop. The objective of the proposed *prediction S-MAC* protocol is minimizing average packet delay compared to the *prolong listen* protocol. Average packet delay is calculated as follows:

Average packet Delay =

$$\left( \frac{\sum_{\forall \text{ packets}} (\text{Arrival time at destination} - \text{Initial time at source})}{\text{Total number of packets}} \right) \text{ in ms}$$

2) *Throughput*: Throughput (often measured in bits or bytes or packets per second) refers to the amount of data successfully transferred from a sender to a receiver in a given time. Many factors affect throughput, including efficiency of collision avoidance, channel utilization, delay and control packet overhead. As with delay, the importance of throughput depends on the application. The proposed prediction S-MAC protocol's objective is to increase throughput compared to other protocols. Throughput is calculated as follows:

$$\text{Throughput} = \frac{\text{Total number of packets (X)}}{\text{Largest arrival time} - \text{Smallest initial time}} \text{ in pkts/sec.}$$

### 3) Average node power consumption:

With large numbers of battery-powered nodes, it is very difficult to change or recharge batteries for these nodes. On many hardware platforms, the radio is a major energy consumer. The energy consumption of the node is measured by multiplying the amount of time that the radio on each node has spent in different modes: sleep, idle, transmitting and receiving by the required power to operate the radio in that mode. The objective of the proposed prediction S-MAC protocol is to minimize the power consumption of each node compared to the prolong listen protocol. Average node power consumption is calculated as follows:

$$\text{Average node power consumption} = \frac{\sum_{\forall \text{ node}} \left( \sum_{\forall \text{ state}} \left( \begin{array}{l} \text{Time spent by the node in a state} \\ * \text{Power consumed in this state} \end{array} \right) \right)}{\text{Number of nodes}}$$

Where state  $\in \{ \text{idle, transmitting, receiving, sleep} \}$

4) *Average node life*: The lifetime  $T_i$  of node  $i$  is defined as the expected time for the energy  $E_i$  to be exhausted, where each node  $i$  has the limited energy  $E_i$  of node  $i$  to be exhausted. The network lifetime  $T$  of the system is defined as the time when the first sensor  $i$  is drained of its energy, that is to say, the system lifetime  $T$  of a sensor network is the minimum lifetime of all nodes of the network,

$$T = \min \{ T_1, T_2, \dots, T_n \} \quad [11].$$

Because the compared protocols have different algorithms, where *prolong listen protocol* has fixed listen periods and the sleep periods are very long. Therefore, calculating node lifetime using real time (in sec) may increase in case of using more sleep time. So it will not a good parameter, therefore instead of using real time to calculate the node lifetime, we will use number of served packets. That is mean, the node lifetime will not be calculated as the time in second the node will go down after, instead it will be calculated as the number of packets the node can serve before going down. So, average node life is calculated as follows:

1. For each node of both the compared protocols, calculate the following:

- Total number of transmitted packets ( $K$ ).
- Total power consumption of transmitted packets ( $P$ ).

- Divide the total power consumption of transmitted packets ( $P$ ) by the total number of transmitted packets ( $K$ ) to get average packet power consumption ( $P_k$ ).

$$\text{Average packet power consumption ( } P_k \text{ )} = \frac{P}{K} \text{ mw/pkt.}$$

2. Using standard maximum battery power consumption of the sensor node ( $P_s$ ) = 2850 mAh \* 3 V = 8610 mwh. (Each node has two AA alkaline batteries) [12].
3. Divide maximum battery power consumption of the sensor node ( $P_s$ ) by the average packet power consumption ( $P_k$ ) to get the average number of packets that each node should transmit before running out of energy ( $T_k$ ).

$$\text{Average number of served packets } T_k = \frac{P_s}{P_k} \text{ pkts}$$

4. Therefore, the average node life in packets is calculated from the following general equation;

$$\text{Average node life in packets} = \sum_{\forall \text{ node}} \left( \frac{\text{Maximum battery power consumption of the node} * \text{Total number of transmitted packets of that node}}{\text{Total power consumption of transmitted packets of that node}} \right)$$

5. Of course, a protocol that transmits a big number of packets before the nodes running out of energy is considered as longer life.

### B. Simulation Parameters

The simulation program used the following values to build and compare the two protocols:

- Number of nodes ( $N$ ) takes the values 10, 20, 30 and 40 nodes consequently.
- Node's range ( $R$ ) is taken as 100 m \* 100 m.
- Number of packets generated at each message is taken as a random number from 1 to 10 packets/node.
- Message length ( $M$ ) is considered as multiple of a unit packet in the number of packets generated at each node.
- History interval count ( $H$ ) is considered as the first 10 listen intervals.
- Minimum number of messages ( $MSG$ ) that is created at each node is equal to 30 messages.
- The radio power consumption taken in receiving, transmitting and sleeping is 45 mw, 60 mw and 90  $\mu$ w respectively. There is no difference between listening and receiving mode [17].
- Average number of packets/node/sec (Data rate step values ( $A$ )) takes the values 20, 40, 60, 80, 100 and 120 pkts/node/sec.
- Time increasing at the source nodes ( $\Delta t$ ) is a random number.
- Value resulted at each data rate step point ( $A$ ) is the average of running the simulated program five times.
- Confidence interval taken is considered as 95%.

$$(\bar{L}_{N_i} \pm m * \frac{S_i}{\sqrt{N}}, m = 1.96).$$

- Total battery power consumption of the sensor node is calculated by multiplying its volt (1.5 V) by capacity (2870 mAh) (each sensor node has two AA alkaline batteries).



- For the *prolong listen protocol*, the following values are taken:
  - Listen interval ( $L$ ) is fixed and equal to 300 msec.
  - Sleep interval ( $S$ ) is fixed and equal to 1000 msec.
  - Prolong listening time ( $P$ ) is equal to 20 msec.
  - Start listen time of nodes ( $ST$ ) is a random number from 1 to 25 msec.

### C. Results

Both the proposed prediction S-MAC protocol and the prolong listen protocol are simulated using "visual C++ programming language". Performance parameters evaluation resulted from the simulation program; average packet delay, throughput, average node power consumption and average node life are computed to evaluate the compared protocols. For each parameter, four figures (from a to d) are used to compare the two protocols changing number of nodes ( $N$ ) from ten to forty nodes by a step of ten. Average number of packets/node/sec (data rate step values) used are; 20, 40, 60, 80, 100 and 120 packets/node/sec.

For the prolong listen protocol; listen time ( $L$ ) is fixed at 300 ms. while sleep time ( $S$ ) is fixed at 1000 ms. Also, prolong listen time ( $P$ ) used is 20 ms. while start listen time of each node is random time varying from 1 to 25 ms. To simulate reality, parameters used in the simulation program are generated randomly and in order to obtain accurate results, each point in the data rate step values is the average of running the simulated program five times.

#### 1) Average packet delay

It is known that real packet delay is the sum of waiting time and transmission time. As shown in figure 11, It is clear that in case of the prolong listen protocol, increasing sleep time leads to higher average packet delay since the time that a packet needs to wait for the node to enter listen mode increases. Also, note that for the two protocols, increasing both average number of packets per node per second (data rate step values) and number of nodes lead to increasing average packet delay as shown in figure 11.

Note that, using the *prolong listen protocol* leads to a higher average packet delay than using *prediction S-MAC protocol*. The results are expected since in case of the *prolong listen protocol*, packets always wait a long time (long sleep intervals) for the node to enter the listen mode. While in case of the *prediction S-MAC protocol*, there is no fixed listen and sleep intervals. However, the nodes transmit only (send/receive) according to the prediction of their listen time, otherwise the nodes go to sleep mode and turn off their radio power until next prediction listen time. This implies to decreasing waiting time compared to the prolong listen protocol. So, the proposed prediction S-MAC protocol decreases average packet delay compared to the prolong listen protocol as shown in figure 11.

#### 2) Throughput

Throughput refers to the amount of data successfully transferred from a sender to a receiver in a given time (often measured in bits or bytes or packets per second). As shown in figure 12, it is clear that in case of the prolong listen protocol; long sleep intervals lead to lower throughput than in the

prediction S-MAC protocol. Since packets in the prolong listen protocol have to wait for a longer time for the nodes to enter their listen interval, so fewer packets are delivered per second (throughput). Of course for the two protocols, in the highest traffic load (increasing both data rates and number of nodes), contention happens at each hop, which can significantly reduce throughput. This leads to semi-straight lines appearing when the number of nodes is equal to forty nodes.

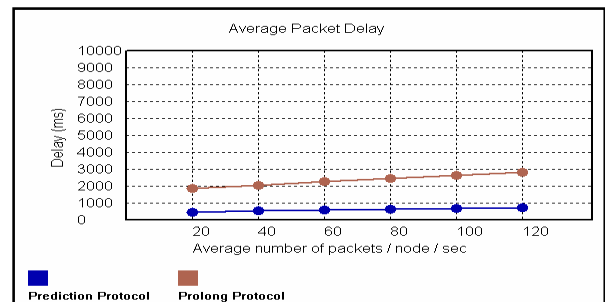


Figure 11.a. Average packet delay at  $N = 10$  nodes.

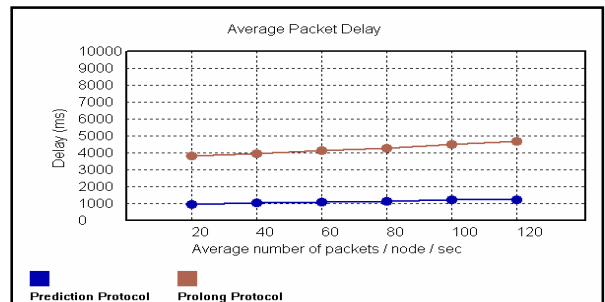


Figure 11.b. Average packet delay at  $N = 20$  nodes.

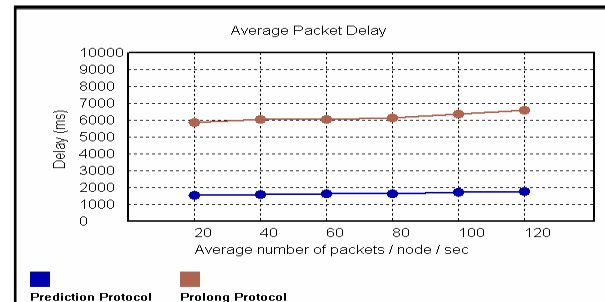


Figure 11.c. Average packet delay at  $N = 30$  nodes.

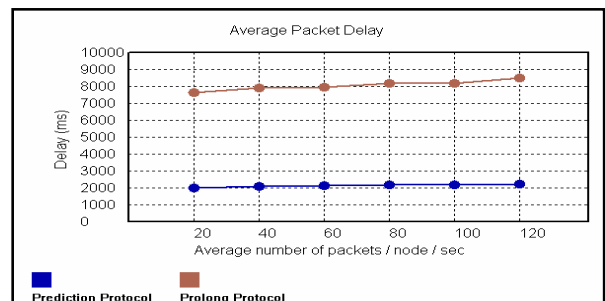


Figure 11.d. Average packet delay at  $N = 40$  nodes.

Figure 11. Average packet delay.

As shown in figure12 using prolong listen protocol leads to lower throughput than using prediction S-MAC protocol. These results are expected since in case of the prolong listen protocol, packets always have to wait for the nodes entering their sleep mode, so number of packets delivered per second (*throughput*) decreases. While using prediction S-MAC protocol improves throughput by a great value since there is no fixed listen & sleep intervals, so packets do not wait long sleep time to be transmitted. Therefore, packets are served in the same listen interval instead of waiting for next listen intervals. Thus, number of packets delivered per second (*throughput*) increases and this leads to increasing throughput compared to the prolong listen protocol as shown in the figures.

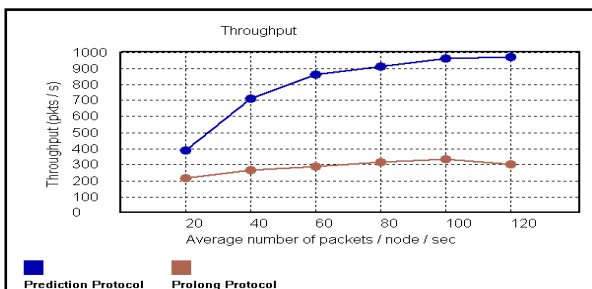


Figure 12.a. Throughput at  $N = 10$  nodes.

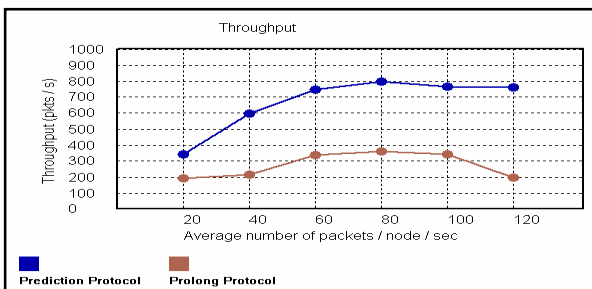


Figure 12.b. Throughput at  $N = 20$  nodes.

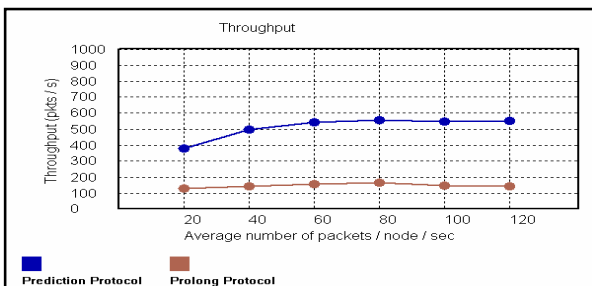


Figure 12.c. Throughput at  $N = 30$  nodes.

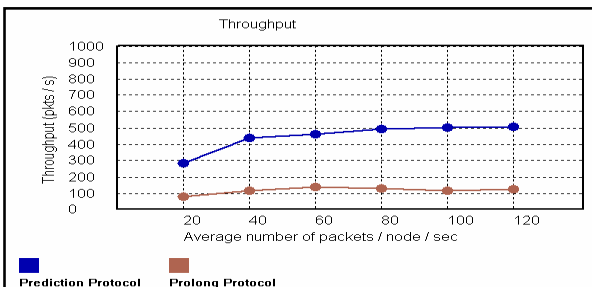


Figure 12.d. Throughput at  $N = 40$  nodes.

Figure12. Throughput

### 3) Average node power consumption

Energy consumption for each node is calculated by multiplying the energy consumed at each mode (*sleep, idle, transmitting and receiving*) by the time that the node has spent in that mode. As shown in figure 13 note that, using the *prolong listen protocol* leads to higher average node power consumption than using the *prediction S-MAC protocol*. Since in case of the *prolong listen protocol* remind that, any node does not send a *RTS* message unless the destination is in listen mode and packets have to wait a long time for the destination node to enter the listen state. In fact, some queued packets may have to wait more than one period if their nodes are serving others, this implies to increasing sleep time and idle time. In addition, some *RTS* messages were sent by the source nodes and may be not answered. These *RTS* messages have to be resent and increase the transmission time. Although not all the next-hop nodes could overhear the *RTR* messages from the transmitting nodes, since they are not at the scheduled listen time or they do not have data packets to send. Therefore, if a node starts a transmission by sending out a *RTR* message during prolong listen time, it might not get a reply. In this case, it just goes back to sleep mode and will try again at the next normal listen time, which increases transmission time and *RTR* messages consume energy. Thus, as sleep time increases, average node power consumption increases.

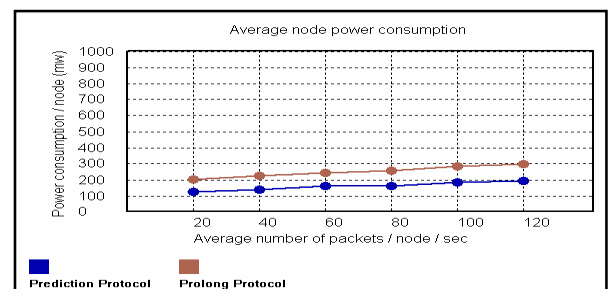


Figure 13.a. Av. node pw. consumption at  $N = 10$  nodes.

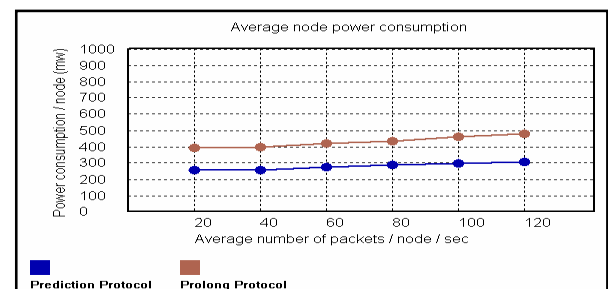


Figure 13.b. Av. node pw. consumption at  $N = 20$  nodes.

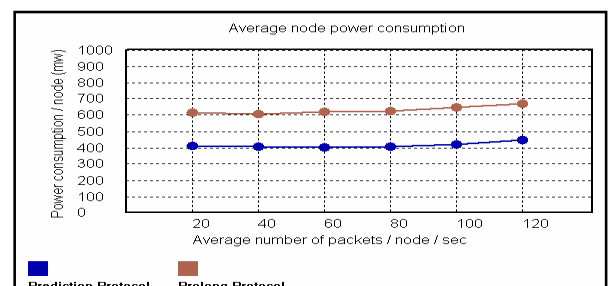


Figure 13.c. Av. node pw. consumption at  $N = 30$  nodes.

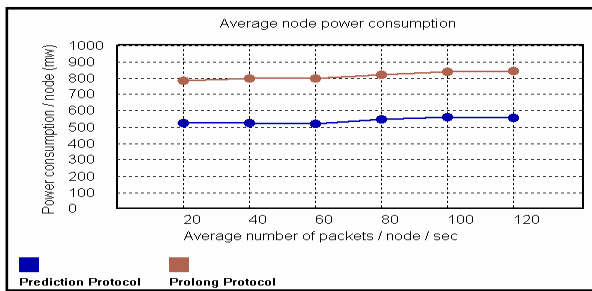


Figure 13.d. Av. node pw. consumption at  $N = 40$  nodes.

While in case of the *prediction S-MAC protocol*, there is no fixed listen and sleep intervals and nodes become active only when transmitting (*send or receive*), otherwise nodes turn off radio power until their next prediction active time. Therefore, packets do not have to wait a long time for the node to enter the active state. Thus idle time is decreased. In addition, there are no *RTR* messages and just small numbers of *RTS* messages is repeated. This leads to lower transmission time than *prolong listen protocol*. As a result, the *prediction S-MAC protocol* decreases the average node power consumption compared to the *prolong listen protocol*. In addition, it is clear that for the two protocols, increasing both average number of packets per node per second (*data rate step values*) and number of nodes imply to increasing average node power consumption as shown in figure 13.

#### 4) Average node life

The lifetime  $T_i$  of node  $i$  is usually defined as the expected time for the energy  $E_i$  to be exhausted, where each node  $i$  has the limited energy  $E_i$  of node  $i$  to be exhausted [11]. Considering the amount of time until the sensor node runs out of energy to refer to the average node life is not fair. So as a good idea instead of using the amount of time to calculate the average node life, the number of packets each node can serve before the node runs out of energy is used to refer to the average node life. That means; a protocol that serves a larger number of packets before the first sensor node die has longer life than the other.

As shown in figure 14, it is clear that the proposed prediction S-MAC protocol serves a bigger number of packets before the nodes exhaust their power compared to the prolong listen protocol. Therefore, the proposed prediction S-MAC protocol has a longer node life than the prolong listen protocol. These results are logic since packets in the prolong listen protocol have to wait for a longer time for the nodes to enter their listen mode, so fewer packets are delivered per second and nodes' battery are exhausted during long waiting time.

While in case of the *prediction S-MAC protocol*, there is no fixed listen and sleep intervals where, nodes active only when transmitting (*send or receive*) and turn off their radio power until next expected working time. In addition, packets do not have to wait a long time for the node to enter the active state and almost wait only if the destination node is busy. Therefore, idle time is decreased and number of served packets are increased before nodes run out of energy and consequently increasing average node life compared to the *prolong listen protocol*. Of course for the two protocols, increasing data rates and increasing number of nodes lead to decreasing average node life.

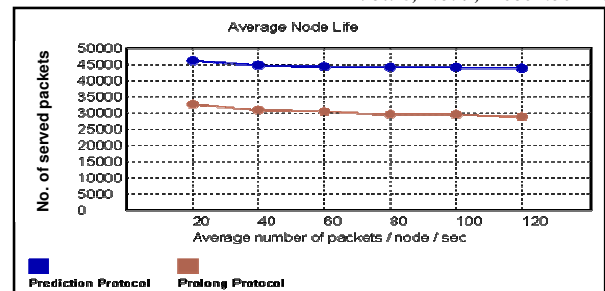


Figure 14.a. Average node life at  $N = 10$  nodes.

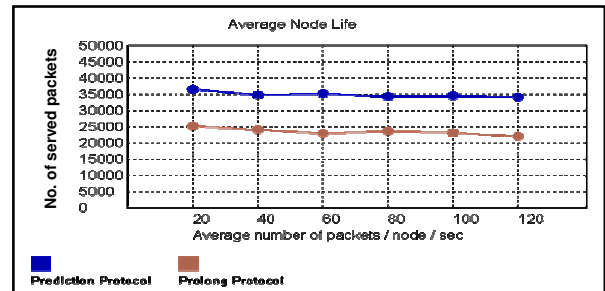


Figure 14.b. Average node life at  $N = 20$  nodes.

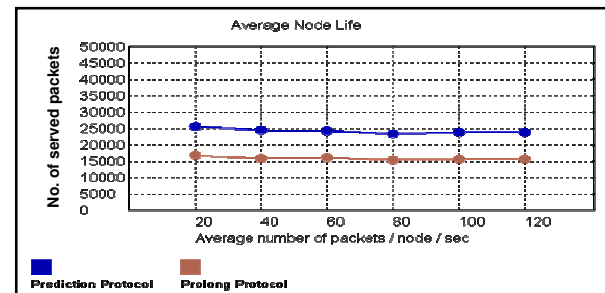


Figure 14.c. Average node life at  $N = 30$  nodes.

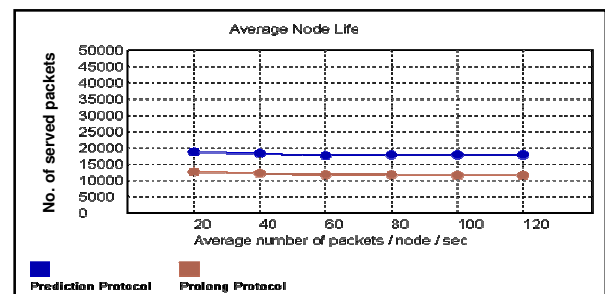


Figure 14.d. Average node life at  $N = 40$  nodes.

Figure 14. Average node life.

*As final words; both the proposed prediction S-MAC protocol and the prolong listen protocol are simulated by using visual C++ programming language. Performance parameters evaluation resulted from the simulation program; average packet delay, throughput, average node power consumption and average node life are evaluated for both the compared protocols. Results illustrate that the proposed prediction S-MAC protocol improves the performance of the network compared to the prolong listen protocol; it leads to lower average packet delay, higher throughput, lower average node power*

**consumption and longer average node life. In addition, for the two protocols; increasing both the average number of packets per node per second (data rate step values) and number of nodes lead to increase both the average packet delay and the average node power consumption while decrease both the throughput and average node life.**

## V. CONCLUSION AND FUTURE TRENDS

In this paper, a new S-MAC protocol named; "*prediction S-MAC protocol*" is proposed to improve the performance of the previous S-MAC protocols. The basic idea of the proposed protocol is to divide the whole time of the node into two successive intervals; working interval (*listen interval*), in which the node is expected to send or receive packets and non-working interval (*sleep interval*), in which the node is not expected to send or receive packets.

Confidence interval method and law of large numbers are used to predict the *working* and *non-working* (*listen/sleep*) intervals based on the last previous  $N$  listen intervals by the ratios 90%, 95%, 99% (*or whatever*). The proposed prediction S-MAC protocol was compared with the prolong listen protocol which was considered as the best protocol of the existing S-MAC protocols. Results proved that the proposed prediction S-MAC protocol increased both throughput and average node life while decreased both delay and average node power consumption compared to the prolong listen protocol.

As a future work, we will try to get a model which gives both estimation and prediction of the future energy consumption in sensor nodes. This model is based on the statistics methods such as Markov chains. If the sensor node can predict its power consumption then it would be better to transmit the predicted energy in the batteries for the path discovery, this will allow also a priori reaction and a possible optimization of the mechanism applied for the minimization of the energy consumption, which depends essentially on the remaining energy in sensors batteries.

## VI. REFERENCES

1. T.S. Rappaport, "*Wireless Communications, Principles and Practice*", Prentice Hall, 2edition, Dec 31, 2001.
2. Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "*A Survey on Sensor Networks*", IEEE Communications Magazine, vol. 50, no.8, pp. 102-114, August 2002.
3. J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan, "*Building efficient wireless sensor networks with low-level naming*", Proceedings of the eighteenth ACM symposium on Operating systems principles, Banff, Alberta, Canada, pp. 146-159, October 2004.
4. Deepak Ganesan, Alberto Cerpa, Wei Ye, Yan Yu, Jerry Zhao and Deborah Estrin, "*Networking Issues in Wireless Sensor Networks*", Journal of parallel and distributed computing, vol.64, no.7, pp.799-814, July 2004.
5. O. Ocakoglu and O. Ercetin, "*Energy Efficient Random Sleep-Awake Schedule Design*", IEEE communication letters, Vol. 10, No.7, JULY 2006.
6. Wei Ye, John Heidemann and Deborah Estrin, "*An energy-efficient MAC protocol for wireless sensor networks*", Proceedings of the IEEE INFOCOM, New York, USA, pp.1567-1576, June 2002.
7. Wei Ye, John Heidemann and Deborah Estrin, "*Medium Access Control with Coordinated, Adaptive Sleeping for wireless sensor networks*", IEEE/ACM transactions on networking, vol. 12, no 3, pp. 493-506, June 2004.
8. Wei Ye and John Heidemann, "*Medium Access Control in Wireless Sensor Networks*", Proceedings of the USC/ISI TECHNICAL REPORT ISI-TR580, October 2003.
9. Imane A. Saroit and Mahmoud A. El-Sakhawy, "*Propose Medium Access Control for Wireless Sensor Networks*", published in the third annual conference (INFOS 2005), faculty of computers and information, Cairo university, 19-22 March 2005. [www.fci-cu.edu.eg/infos2005](http://www.fci-cu.edu.eg/infos2005).
10. Douglas G Altman, David Machin, Trevor N Bryant and Martin J Gardener, "*Statistics with confidence*", 2nd edition.
11. Yunxia Chen and Qing Zhao, "*Maximizing the Lifetime of Sensor Network Using Local Information on Channel State and Residual Energy*", Conference on Information Sciences and Systems, The Johns Hopkins University, 2008.
12. Michael Day, "*Using power solutions to extend battery life in applications*", Analog Applications Journal, 2009.
13. Yuan Li, Wei Ye, John Heidemann, and Rohit Kulkarni, "*Design and Evaluation of Network Reconfiguration Protocols for Mostly-Off Sensor Networks*" Information Sciences Institute, University of Southern California, 2007.

# Simulation of Grover's Algorithm Quantum Search in a Classical Computer

A. B. Mutiara

Faculty of Computer Science and Information Technology  
Gunadarma University  
Jl. Margonda Raya No.100, Depok 16424, Indonesia  
amutiara@staff.gunadarma.ac.id

R.Refianti

Faculty of Computer Science and Information Technology  
Gunadarma University  
Jl. Margonda Raya No.100, Depok 16424, Indonesia  
rina@staff.gunadarma.ac.id

**Abstract—** The rapid progress of computer science has been accompanied by a corresponding evolution of computation, from classical computation to quantum computation. As quantum computing is on its way to becoming an established discipline of computing science, much effort is being put into the development of new quantum algorithms. One of quantum algorithms is Grover's algorithm, which is used for searching an element in an unstructured list of  $N$  elements with quadratic speed-up over classical algorithms. In this work, Quantum Computer Language (QCL) is used to make a Grover's quantum search simulation in a classical computer document.

**Keywords:** Grover's Algorithm, Quantum Computer Language, Hadamard-Transform

## I. INTRODUCTION

The rapid progress of computer science has been accompanied by a corresponding evolution of computation, from classical computation to quantum computation. In classical computation, computer memory made up of bits, where each bit represents either a one or a zero. In quantum computation, there are some quantum mechanical phenomena, such as superposition and entanglement, to perform operations on data.

Instead of using bits, quantum computation uses qubits (quantum bits). A single qubit can represent a one, a zero, or both at the same time, which is called superposition. Because of this ability, quantum computation can perform many tasks simultaneously, faster than classical computing. There is also another phenomenon in quantum computation which is called entanglement. If two qubits get an outside force, then those qubits can be entangled condition. It means that, even the distance of both qubits is far, treating one of them will affect the other qubit too. For example, there are two entangled qubits, and one of them has spin up (we know it after done a measurement). Then without have to measure it, we can directly know that the other qubit has spin down. Because of this ability, communication in quantum computation can reach a very high speed because information can be transferred instantly, very fast like it overmatches the speed of light.

As quantum computing is on its way to becoming an established discipline of computing science, much effort is being put into the development of new quantum algorithms.

One of quantum algorithms is Grover algorithm, which is used for searching an element in an unstructured list of  $N$  elements with quadratic speed-up over classical algorithms. Today, there are some quantum programming languages which can be used to simulate quantum mechanical and quantum algorithm without having a real quantum computer. In this work, Quantum Computer Language (QCL) will be used to make a Grover's quantum search simulation in a classical computer.

This research is related to an invention of a quantum search algorithm by Lov K. Grover [1]. His invention presents an algorithm, which is known as Grover algorithm that is significantly faster than any classical algorithm can be. This quantum search algorithm can search for an element in an unsorted database containing  $N$  elements only in  $O(\sqrt{N})$  steps, while in the models of classical computation, searching an unsorted database cannot be done in less than linear time (so merely searching through every item is optimal), which will be done in  $O(N)$  steps. Also, this research is related with Paramita et al work, where their paper presented a pseudo code for better understanding about Grover algorithm, and Freddy P. Zen et. al. work [2], who provide an example simulation of Grover algorithm in their paper. This research will also try to simulate Grover algorithm in a classical computer using one of quantum programming languages, Quantum Computer Language (QCL) [3,4].

In practice, this research can be used as the fastest known method or solution for searching an element in an unsorted database containing  $N$  elements. By using the method in this research, the searching process can speed-up quadratically over classical algorithms.

Considered points in this work are:

1. Is it possible to simulate Grover algorithm in a classical computer?
2. How many qubits and iterations the program needed to search an element?
3. How minimum and maximum the size of elements in the database that the program can hold?

The objective of this work is to make a simulation of Grover algorithm using Quantum Computer Language (QCL), to know how many qubits and iterations needed for the

searching process, and to know how minimum and maximum the size of elements in the database that can be hold by the program.

This paper concerns on simulating Grover algorithm in a classical computer using Quantum Computer Language (QCL). The program can search a desired element in an unsorted database of N elements.

This work begins with designing pseudo code and flowchart for Grover algorithm. Then, the design will be implemented by using Quantum Computer Language. After that, there will be several test to know how many qubits and iterations needed for the searching process, also to know how minimum and maximum the size of elements in the database that can be hold by the program.

## II. LITERATURE REVIEW

Computer science has grown faster, made an evolution in computation. Research has already begun on what comes after our current computing revolution. This research has discovered the possibility for an entirely new type of computer, one that operates according to the laws of quantum physics - a quantum computer.

### A. Way to Quantum Computation

Quantum computers were first proposed in the 1970s and 1980s by theorists such as Richard Feynman, Paul Benioff, and David Deutsch. At those times, many scientists doubted that they could ever be made practical. Richard Feynman was the first to suggest, in a talk in 1981, that quantum-mechanical systems might be more powerful than classical computers. In this lecture [5], reproduced in the International Journal of Theoretical Physics in 1982, Feynman asked what kind of computer could simulate physics and then argued that only a quantum computer could simulate quantum physics efficiently. He focused on quantum physics rather than classical physics. He said that nature isn't classical, and if we want to make a simulation of nature, we'd better make it quantum mechanical, because it does not look so easy. Around the same time, in a paper titled "Quantum mechanical models of Turing machines that dissipate no energy" [6] and related articles, Paul Benioff demonstrated that quantum-mechanical systems could model Turing machines. In other words, he proved that quantum computation is at least as powerful as classical computation. But is quantum computation more powerful than classical computation? David Deutsch explored this question and more in his 1985 paper "Quantum theory, the Church-Turing principle and the universal quantum computer" [7]. First, he introduced quantum counterparts to both the Turing machine and the universal Turing machine. He then demonstrated that the universal quantum computer can do things that the universal Turing machine cannot, including generate genuinely random numbers, perform some parallel calculations in a single register, and perfectly simulate physical systems with finite dimensional state spaces. In 1989, in "Quantum computational networks"[8], Deutsch described a second model for quantum computation: quantum circuits. He demonstrated that quantum

gates can be combined to achieve quantum computation in the same way that Boolean gates can be combined to achieve classical computation. He then showed that quantum circuits can compute anything that the universal quantum computer can compute, and vice versa.

### B. Quantum Computer Development

Quantum computers could one day replace silicon chips, just like the transistor once replaced the vacuum tube. But for now, the technology required to develop such a quantum computer is beyond our reach. Most research in quantum computing is still very theoretical.

The most advanced quantum computers have not gone beyond manipulating more than 16 qubits, meaning that they are a far cry from practical application. However, the potential remains that quantum computers one day could perform, quickly and easily, calculations that are incredibly time-consuming on conventional computers. Several key advancements have been made in quantum computing in the last few years. Let's look at a few of the quantum computers that have been developed.

- In 1998, Los Alamos and MIT researchers managed to spread a single qubit across three nuclear spins in each molecule of a liquid solution of alanine (an amino acid used to analyze quantum state decay) or trichloroethylene (a chlorinated hydrocarbon used for quantum error correction) molecules. Spreading out the qubit made it harder to corrupt, allowing researchers to use entanglement to study interactions between states as an indirect method for analyzing the quantum information.
- In March 2000, scientists at Los Alamos National Laboratory announced the development of a 7-qubit quantum computer within a single drop of liquid. The quantum computer uses nuclear magnetic resonance (NMR) to manipulate particles in the atomic nuclei of molecules of trans-crotonic acid, a simple fluid consisting of molecules made up of six hydrogen and four carbon atoms. The NMR is used to apply electromagnetic pulses, which force the particles to line up. These particles in positions parallel or counter to the magnetic field allow the quantum computer to mimic the information-encoding of bits in digital computers. Researchers at IBM-Almaden Research Center developed what they claimed was the most advanced quantum computer to date in August. The 5-qubit quantum computer was designed to allow the nuclei of five fluorine atoms to interact with each other as qubits, be programmed by radio frequency pulses and be detected by NMR instruments similar to those used in hospitals (see How Magnetic Resonance Imaging Works for details). Led by Dr. Isaac Chuang, the IBM team was able to solve in one step a mathematical problem that would take conventional computers repeated cycles. The problem, called order-finding, involves finding the period of a particular function, a typical aspect of many mathematical problems involved in cryptography.
- In 2005, the Institute of Quantum Optics and Quantum Information at the University of Innsbruck announced that



scientists had created the first qubyte, or series of 8 qubits, using ion traps.

- In 2006, Scientists in Waterloo and Massachusetts devised methods for quantum control on a 12-qubit system. Quantum control becomes more complex as systems employ more qubits.
- In 2007, Canadian startup company D-Wave demonstrated a 16-qubit quantum computer. The computer solved a sudoku puzzle and other pattern matching problems. The company claims it will produce practical systems by 2008. Skeptics believe practical quantum computers are still a decade away that the system D-Wave has created isn't scaleable, and that many of the claims on D-Wave's Web site are simply impossible (or at least impossible to know for certain given our understanding of quantum mechanics).

If functional quantum computers can be built, they will be valuable in factoring large numbers, and therefore extremely useful for decoding and encoding secret information. If one were to be built today, no information on the Internet would be safe. Our current methods of encryption are simple compared to the complicated methods possible in quantum computers. Quantum computers could also be used to search large databases in a fraction of the time that it would take a conventional computer. Other applications could include using quantum computers to study quantum mechanics, or even to design other quantum computers.

But quantum computing is still in its early stages of development, and many computer scientists believe the technology needed to create a practical quantum computer is years away. Quantum computers must have at least several dozen qubits to be able to solve real-world problems, and thus serve as a viable computing method.

### C. Superposition

Superposition is the fundamental law of quantum mechanics. It defines the collection of all possible states that an object can have. Superposition means a system can be in two or more of its states simultaneously. For example a single particle can be traveling along two different paths at once.

The principle of superposition states that if the world can be in any configuration, any possible arrangement of particles or fields, and if the world could also be in another configuration, then the world can also be in a state which is a superposition of the two, where the amount of each configuration that is in the superposition is specified by a complex number.

For example, if a particle can be in position A and position B, it can also be in a state where it is an amount "3i/5" in position A and an amount "4/5" in position B. To write this, physicists usually say:

$$|\psi\rangle = \frac{3}{5}i|A\rangle + \frac{4}{5}|B\rangle \quad (1)$$

In the description, only the relative size of the different components matter and their angle to each other are on the complex plane. This is usually stated by declaring that two states which are a multiple of one another are the same as far as the description of the situation is concerned.

$$|\psi\rangle \approx \alpha|\psi\rangle \quad (2)$$

The fundamental dynamical law of quantum mechanics is that the evolution is linear, meaning that if the state A turns into A' and B turns into B' after 10 seconds, then after 10 seconds the superposition  $\psi$  turns into a mixture of A' and B' with the same coefficients as A' and B'.

*Example:* A particle can have any position, so that there are different states which have any value of the position  $x$ . These are written:

$$|x\rangle$$

The principle of superposition guarantees that there are states which are arbitrary superpositions of all the positions with complex coefficients:

$$\sum_x \psi(x)|x\rangle$$

This sum is defined only if the index  $x$  is discrete. If the index is over  $R$ , then the sum is not defined and is replaced by an integral instead. The quantity  $\psi(x)$  is called the wavefunction of the particle.

If a particle can have some discrete orientations of the spin, say the spin can be aligned with the  $z$ -axis  $|+\rangle$  or against it  $|-\rangle$ , then the particle can have any state of the form:

$$C_1|+\rangle + C_2|-\rangle$$

If the particle has both position and spin, the state is a superposition of all possibilities for both:

$$\sum_x \psi_+(x)|x,+\rangle + \psi_-(x)|x,-\rangle$$

The configuration space of a quantum mechanical system cannot be worked out without some physical knowledge. The input is usually the allowed different classical configurations, but without the duplication of including both position and momentum.

A pair of particles can be in any combination of pairs of positions. A state where one particle is at position  $x$  and the other is at position  $y$  is written  $|x,y\rangle$ . The most general state is a superposition of the possibilities:

$$\sum_{xy} A(x, y) |x, y\rangle$$

The description of the two particles is much larger than the description of one particle; it is a function in twice the number of dimensions. This is also true in probability, when the statistics of two random things are correlated. If two particles are uncorrelated, the probability distribution for their joint position  $P(x, y)$  is a product of the probability of finding one at one position and the other at the other position:

$$P(x, y) = P_x(x)P_y(y) \quad (3)$$

In quantum mechanics, two particles can be in special states where the amplitudes of their position are uncorrelated. For quantum amplitudes, the word entanglement replaces the word correlation, but the analogy is exact. A disentangled wavefunction has the form:

$$A(x, y) = \psi_x(x)\psi_y(y) \quad (4)$$

while an entangled wavefunction does not have this form. Like correlation in probability, there are many more entangled states than disentangled ones. For instance, when two particles which start out with an equal amplitude to be anywhere in a box have a strong attraction and a way to dissipate energy, they can easily come together to make a bound state. The bound state still has an equal probability to be anywhere, so that each particle is equally likely to be everywhere, but the two particles will become entangled so that wherever one particle is, the other is too.

#### D. Entanglement

Quantum entanglement, also called the quantum non-local connection, is a property of a quantum mechanical state of a system of two or more objects in which the quantum states of the constituting objects are linked together so that one object can no longer be adequately described without full mention of its counterpart - even if the individual objects are spatially separated in a space-like manner. The property of entanglement was understood in the early days of quantum theory, although not by that name. Quantum entanglement is at the heart of the EPR paradox developed in 1935. This interconnection leads to non-classical correlations between observable physical properties of remote systems, often referred to as nonlocal correlations.

Quantum mechanics holds that observable, for example, spin are indeterminate until such time as some physical intervention is made to measure the observable of the object in question. In the singlet state of two spins it is equally likely that any given particle will be observed to be spin-up as that it will be spin-down. Measuring any number of particles will result in an unpredictable series of measures that will tend more and more closely to half up and half down. However, if this experiment is done with entangled particles the results are quite different. For example, when two members of an entangled pair are measured, their spin measurement results will be correlated.

Two (out of infinitely many) possibilities are that the spins will be found to always have opposite spins (in the spin anti-correlated case), or that they will always have the same spin (in the spin correlated case). Measuring one member of the pair therefore tells you what spin the other member would have if it were also measured. The distance between the two particles is irrelevant.

Theories involving 'hidden variables' have been proposed in order to explain this result; these hidden variables account for the spin of each particle, and are determined when the entangled pair is created. It may appear then that the hidden variables must be in communication no matter how far apart the particles are that the hidden variable describing one particle must be able to change instantly when the other is measured. If the hidden variables stop interacting when they are far apart, the statistics of multiple measurements must obey an inequality (called Bell's inequality), which is, however, violated - both by quantum mechanical theory and in experiments.

When pairs of particles are generated by the decay of other particles, naturally or through induced collision, these pairs may be termed "entangled", in that such pairs often necessarily have linked and opposite qualities, i.e. of spin or charge. The assumption that measurement in effect "creates" the state of the measured quality goes back to the arguments of, among others: Schroedinger, and Einstein, Podolsky, and Rosen concerning Heisenberg's uncertainty principle and its relation to observation (see also the Copenhagen interpretation). The analysis of entangled particles by means of Bell's theorem, can lead to an impression of non-locality (that is, that there exists a connection between the members of such a pair that defies both classical and relativistic concepts of space and time). This is reasonable if it is assumed that each particle departs the scene of the pair's creation in an ambiguous state (as per a possible interpretation of Heisenberg). In such a case, for a given measurement either outcome remains a possibility; only measurement itself would precipitate a distinct value. On the other hand, if each particle departs the scene of its "entangled creation" with properties that would unambiguously determine the value of the quality to be subsequently measured, then a postulated instantaneous transmission of information across space and time would not be required to account for the result. The Bohm interpretation postulates that a guide wave exists connecting what are perceived as individual particles such that the supposed hidden variables are actually the particles themselves existing as functions of that wave.

Observation of wavefunction collapse can lead to the impression that measurements performed on one system instantaneously influence other systems entangled with the measured system, even when far apart. Yet another interpretation of this phenomenon is that quantum entanglement does not necessarily enable the transmission of classical information faster than the speed of light because a classical information channel is required to complete the process.

#### E. Hadamard Transform

The Hadamard transform (also known as the Walsh-Hadamard-transform, Hadamard-Rademacher-Walsh-trans-

form, Walsh-transform, or Walsh-Fourier transform) is an example of a generalized class of Fourier transforms. It is named for the French mathematician Jacques Solomon Hadamard, the German-American mathematician Hans Adolph Rademacher, and the American mathematician Joseph Leonard Walsh. It performs an orthogonal, symmetric, involutorial, linear operation on  $2m$  real numbers (or complex numbers, although the Hadamard matrices themselves are purely real).

The Hadamard transform can be regarded as being built out of size-2 discrete Fourier transforms (DFTs), and is in fact equivalent to a multidimensional DFT of size  $2 \times 2 \cdots 2 \times 2$ . It decomposes an arbitrary input vector into a superposition of Walsh functions.

The Hadamard transform  $H_m$  is a  $2m \times 2m$  matrix, the Hadamard matrix (scaled by a normalization factor), that transforms  $2m$  real numbers  $x$  into  $2m$  real numbers  $x_k$ . The Hadamard transform can be defined in two ways: recursively, or by using the binary (base-2) representation of the indices  $n$  and  $k$ .

Recursively, we define the  $1 \times 1$  Hadamard transform  $H_0$  by the identity  $H_0 = 1$ , and then define  $H_m$  for  $m > 0$  by:

$$H_m = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{m-1} & H_{m-1} \\ H_{m-1} & -H_{m-1} \end{pmatrix} \quad (5)$$

where the  $1/\sqrt{2}$  is a normalization that is sometimes omitted. Thus, other than this normalization factor, the Hadamard matrices are made up entirely of 1 and -1.

Equivalently, we can define the Hadamard matrix by its  $(k, n)$ -th entry by writing

$$k = k_{m-1}2^{m-1} + k_{m-2}2^{m-2} + \cdots + k_12 + k_0, \quad (6)$$

and

$$n = n_{m-1}2^{m-1} + n_{m-2}2^{m-2} + \cdots + n_12 + n_0, \quad (7)$$

where the  $k_j$  and  $n_j$  are the binary digits (0 or 1) of  $n$  and  $k$ , respectively. In this case, we have:

$$(H_m)_{k,n} = \frac{1}{2^{m/2}} (-1)^{\sum_j k_j n_j}. \quad (8)$$

This is exactly the multidimensional  $2 \times 2 \times \cdots \times 2 \times 2$  DFT, normalized to be unitary, if the inputs and outputs are regarded as multidimensional arrays indexed by the  $n_j$  and  $k_j$ , respectively. Some examples of the Hadamard matrices follow.

$$H_0 = +1 \quad (9)$$

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (10)$$

This  $H_1$  is precisely the size-2 DFT. It can also be regarded as the Fourier transform on the two-element additive group of  $\mathbb{Z}/(2)$ .

$$H_2 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (11)$$

$$H_3 = \frac{1}{2^{3/2}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \quad (12)$$

$$(H_n)_{i,j} = \frac{1}{2^{n/2}} (-1)^{i \cdot j}. \quad (13)$$

where  $i \cdot j$  is the bitwise dot product of the binary representations of the numbers  $i$  and  $j$ . For example,  $H_{32} = (-1)^{3 \cdot 2} = (-1)^{(1,1) \cdot (1,0)} = (-1)^{1+0} = (-1)^1 = -1$ , agreeing with the above (ignoring the overall constant). Note that the first row, first column of the matrix is denoted by  $H_{00}$ . The rows of the Hadamard matrices are the Walsh functions.

In quantum information processing the Hadamard transformation, more often called Hadamard gate, is a one-qubit rotation, mapping the qubit-basis states  $|0\rangle$  and  $|1\rangle$  to two superposition states with equal weight of the computational basis states  $|0\rangle$  and  $|1\rangle$ . Usually the phases are chosen so that we have

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}} \langle 0| + \frac{|0\rangle - |1\rangle}{\sqrt{2}} \langle 1|$$

in Dirac notation. This corresponds to the transformation matrix

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (14)$$

in the  $|0\rangle, |1\rangle$  basis.

Many quantum algorithms use the Hadamard transform as an initial step, since it maps  $n$  qubits initialized with  $|0\rangle$  to a superposition of all  $2^n$  orthogonal states in the  $|0\rangle, |1\rangle$  basis with equal weight.  
Hadamard gate operations:

$$H|1\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle. \quad (15)$$

$$H|0\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle. \quad (16)$$

$$H\left(\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle\right) = \frac{1}{2}(|0\rangle + |1\rangle) - \frac{1}{2}(|0\rangle - |1\rangle) = |1\rangle; \quad (17)$$

$$\begin{aligned} H\left(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle\right) &= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) + \frac{1}{\sqrt{2}}\left(-\frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle\right) \\ &= |0\rangle. \end{aligned} \quad (18)$$

#### F. Grover's Quantum Search Algorithm

One of the most celebrated achievement of quantum computation is Lov Grover's quantum search algorithm (known as Grover's algorithm), which was invented in 1996. Grover's algorithm is a quantum algorithm for searching an unsorted database with  $N$  entries in  $O(N^{1/2})$  time and using  $O(\log N)$  storage space.

In models of classical computation, searching an unsorted database cannot be done in less than linear time (so merely searching through every item is optimal). Grover's algorithm illustrates that in the quantum model searching can be done faster than this; in fact its time complexity  $O(N^{1/2})$  is asymptotically the fastest possible for searching an unsorted database in the quantum model. It provides a quadratic speedup.

There are already related works about Grover algorithm, such as done by S. Paramita et.al., Matthew Whitehead, Ahmed Younes, and C. Lavor et.al.. S. Paramita et.al. wrote a pseudo code for Grover algorithm in their paper [9]. They also gave the example of Grover's implementation using their pseudo code. Grover's algorithm can be combined with another search algorithm. Matthew Whitehead's paper [10] shows how Grover's quantum search may be used to improve the effectiveness of traditional genetic search on a classical computer. He uses repeated applications of Grover's Algorithm to get a variety of decent chromosomes that will then be used to form a starting population for classical genetic search. He also provides the pseudo code for the modified genetic search, which is a combination between Grover's quantum search and standard genetic search. Another work related to Grover algorithm is done by Ahmed Younes. In his paper [11], he

described the performance of Grover's algorithm. Also, C. Lavor et al wrote a review about Grover algorithm by means of a detailed geometrical interpretation and a worked out example. Some basic concepts of Quantum Mechanics and quantum circuits are also reviewed.

### III. DESIGN AND IMPLEMENTATION

Many problems in classical computer science can be reformulated as searching a list for a unique element which matches some predefined condition. If no additional knowledge about the search-condition  $C$  is available, the best classical algorithm is a brute-force search i.e. the elements are sequentially tested against  $C$  and as soon as an element matches the condition, the algorithm terminates. For a list of  $N$  elements, this requires an average of  $N/2$  comparisons. By taking advantage of quantum parallelism and interference, Grover found a quantum algorithm [1] which can find the matching element in only  $O(\sqrt{N})$  steps.

In this thesis, Grover algorithm and its implementation will be explained process by process. The algorithm consists of two parts: (1) Input and initialization (2) Main loop. Each of the parts will be explained and implemented one by one below.

#### A. Input and Initialization

##### 1) Input

This simulation needs to know what number it should search, so user will be prompted to input a round number (integer). The implementation of this input process can be seen below.

```
input "Enter an integer that that will be  
find:" ,bil;
```

In the code implementation above we can see that bil is a variable that is used to store the round number.

##### 2) Initialization

Initialization is a process to initiate variables and qubit registers needed in the simulation.

The most important variables that we have to initiate are the number of qubits and the number of iterations needed. Assume that the number of qubits is called sumqubit, and the number of iterations is called iteration.

To calculate the number of qubits needed, we can use this formula:

$$sumqubit = \lfloor (\log_2 bil) + 1 \rfloor \quad (19)$$

To calculate the number of iterations needed, we can use this formula:

$$iteration = \lceil \pi/8 * \sqrt{2^{sumqubit}} \rceil \quad (20)$$

Then, after the value of both sumqubit and iteration are known, another important step to do is to set up the registers for each qubits. Also, some variables need to be listed to for common process; looping, storing result, etc. The code implementation for initialization can be seen below.

```
int sumqubit = floor(log(bil,2))+1;
int iteration = ceil(pi/8*sqrt
2^sumqubit));
int rmeasurement;
int i;
qureg q[sumqubit];
qureg f[1];
print "Number of qubit
are used:",sumqubit;
print "Number of iteration
are needed:",iteration;
print "start searching process..";
```

### B. Main Loop

Main loop is the main process to begin searching. The steps to do in the main loop are:

1. Reset all qubits to  $|0\rangle$  and apply the Hadamard transform to each of them.
2. Repeat the following operation as much as the number of iterations needed (see the initialization part):
  - Rotate the marked state by a phase of  $\pi$  radians ( $I_f^\pi$ ). A query function needs to be applied. The query function is needed to flip the variable  $f$  if  $x$  (the qubits) is equal to 1111...
  - Apply a phase process between  $\pi$  and  $f$ .
  - Undo the query function.
  - Apply a diffusion function. The process are apply Hadamard transform, invert  $q$ , then apply a phase process between  $\pi$  and  $q$  (rotate if  $q=1111..$ ). After that, undo the invert process and undo Hadamard transform.
  - Do an oracle function by measure the quantum register that has been found, then compare the result to the input.

This iterations must be repeated again if the measurement result does not match with the wanted number.

The code implementation of the main loop including the functions in it can be seen below.

```
{
  reset;
  H(q);
  for i= 1 to iteration {
    print "Iteration",i;
    query(q,f,bil);
    CPhase(pi,f);
    !query(q,f,bil);
    diffuse(q);
  }
```

```
    oracle(q,rmeasurement,bil);
  } until rmeasurement==bil;
  reset;
}
```

### Query Procedure

```
procedure query(qureg x,quvoid f,int bil)
{
  int i;
  for i=0 to #x-1 {
    if not bit(bil,i)
      {Not(x[i]);}
  }
  CNot(f,x);
  for i=0 to #x-1 {
    if not bit(bil,i)
      {!Not(x[i]);}
  }
}
```

### Diffuse Procedure

```
procedure diffuse(qureg q)
{
  H(q);
  Not(q);
  CPhase(pi,q);
  !Not(q);
  !H(q);
}
```

### Oracle Procedure

This procedure is for checking whether the measurement result is match with the wanted number or not. In general, the oracle function can be formulated as below.

$$f(x) = \begin{cases} 1 & \text{if } x = x_0 \\ 0 & \text{if } x \neq x_0 \end{cases}$$

$x$  is the indexes in the database, and  $x_0$  is the wanted index. Back to the simulation, before we implement the oracle, we need to do a measurement to check if the number that been found is already matched with the wanted number. The code implementation can be seen below.

```
procedure oracle(qureg q,int hasil-
measurement,bil)
{
  measure q,rmeasurement;
  if rmeasurement==bil {
    print "result of measurement:",
    rmeasurement;
    print "has equaled with the
    searched number...";
  }
  else {
    print "result of measurement:",
```

```
rmeasurement;  
print "not equaled with the  
searched number...";  
}  
}
```

#### IV. RESULTS AND DISCUSSION

The grover's quantum search simulation can be running from Linux's terminal, by going to the directory where the file is put in then typing "qcl -i -b32 SimulasiGrover.qcl". This command will start QCL then run a file named SimulasiGrover.qcl, and providing all qubits that QCL has (32 qubits).

To discuss the results of the program, table I. containing ten outputs from grover's quantum search simulation program is provided.

TABLE I. OUTPUTS FROM THE PROGRAM

Input	Qubits	Iterations	List of Measured Number	Total Iterations
10	4	2	10	2
30	5	3	30	3
175	8	7	175	7
500	9	9	373 - 500	18
1000	10	13	327 - 1000	26
1676	11	18	1676	18
2000	11	18	1645 - 1497 -1493 - 703 - 2000	90
2200	12	26	3765 - 2349 - 2200	78
8111	13	36	8111	36
9999	14	54	9999	54

In table I, column "Input" is for the number that the user wants to find. Column "Qubits" is the total of qubits needed to search the number. Column "Iterations" is the total iterations needed to find one number to be measuring. Column "List of Measured Numbers" is the list of numbers that are found and get measured until the number is same to the input. Column "Total Iterations" is the total of iterations needed to find the correct number. The value of this column is the multiplication of the value in column "Iterations" and the amount of numbers in column "List of Measured Number".

From the table, we can see that the number of qubits and the number of iterations needed are depend on the value of the number that user wants to find. If the number is bigger, so will the qubits and the iterations be. Sometimes, the number that the program found is not matched with the input. If this condition is happen, the program will do the iterations again until the number is matched with the input. But even the program do the iterations more than one round, the total iterations is never exceed the value of the input. We can see this from the table in the column "Total Iterations". But this Grover's quantum search simulation has a limitation; the maximum qubits that the program can use is only 32 qubits (QCL limitation). For the possible real implementation, grover algorithm can be used for searching a record in database and improving the traditional genetic search.

#### V. CONCLUDING REMARKS

Using QCL, a Grover's quantum search simulation has been made. It is performed without using a quantum computer, but using a classic computer. To search an element, the program needs to use qubit instead of bit. The number of qubits needed is depend on the value of the number that we want to find. The bigger the value of the number, the bigger qubits needed. It goes the same with the number of iterations needed. The minimum value for the number is 1, and the maximum value is depending on the qubits needed. At this far, the program has been tested to search number till 9999. This Grover's quantum search is just a simulation to simulate the algorithm, not a real quantum searching program that can be implemented on the real database.

This Grover's quantum search is just a simulation of quantum search in a classic computer. That is some possible works for the future related to Grover algorithm. Some of them is implementing Grover algorithm in a real database using quantum computer, but in this case, the database must be converted in to quantum states which is probably the most difficult thing to do. Another possible work is improving the traditional genetic search by combine it with Grover algorithm.

#### ACKNOWLEDGMENT

A.B.M. and R.R. gratefully acknowledge financial support of the Gunadarma Education Foundation during the research.

#### REFERENCES

- [1] L.K. Grover, "A fast quantum mechanical algorithm for database search", Proceeding of the 28th Annual ACM Symposium on Theory of Computing, 1996, arXiv:quant-ph/9605043v3
- [2] F.P. Zen, A.N. Atmaja, and S.-Sigit, "Pencarian data dengan indeks tak-terurut menggunakan algoritma kuantum", Indonesian Journal of Physics, no.4, October 2003.
- [3] B. Oemer, S.Doiz, and D.K. Svozil, "Simulation of quantum computers", 1996.
- [4] B. Oemer, "Structured quantum programming", May 2003, [http://tph.tuwien.ac.at/\\$~Soemer/](http://tph.tuwien.ac.at/$~Soemer/).
- [5] R.P. Feynmann, "Simulating physics with computers", International Journal of Theoretical Physics, vol.-21, no. 6-7, pp. 467-488, 1982.
- [6] P. Benioff, "Quantum mechanical models of turing machines that dissipate no energy", Physical Review Letters, vol.-48, pp. 1581-1585, June 1982.
- [7] D. Deutsch, "Quantum theory, the Church-Turing principle and the universal quantum computer", Proceedings of the Royal Society of London Series A, vol. A400, pp. 97-117, 1985.
- [8] D. Deutsch, "Quantum computational networks", Proceedings of the Royal Society of London, vol. A425, pp. 73-90, 1989.
- [9] Paramita, R.E. Prihandini, and M.-K. Sulaeman, "Algoritma pencarian dalam daftar tak terurut pada komputasi kuantum (algoritma grover)", 2006, <http://www.informatika.org/>.
- [10] M. Whitehead, "Using grover's algorithm for genetic search", December 2005, [https://www.cs.indiana.edu/\\$~\\$mewhite/files/quantum\\_genetic\\_search.pdf](https://www.cs.indiana.edu/$~$mewhite/files/quantum_genetic_search.pdf).
- [11] A. Younes, "Strength and weakness in grover's quantum search algorithm", November 2008, arXiv:quant-ph/0811.4481v1.

#### AUTHORS PROFILE

A.B. Mutiara is a Professor of Computer Science at Faculty of Computer Science and Information Technology, Gunadarma University.



R. Refianti is a Ph.D-Student at Faculty of Computer Science and Information Technology, Gunadarma University.

## APPENDIX

### Listing Program

```
procedure query(quireg x,quvoid f,int bil)
{
    int i;
    for i=0 to #x-1 {      // x -> NOT
        // (x XOR bil)
        if not bit(bil,i)
            {Not(x[i]);}
    }
    CNot(f,x);              // flip f
    // if x=1111..
    for i=0 to #x-1 {      // x <- NOT
        // (x XOR bil)
        if not bit(bil,i)
            {!Not(x[i]);}
    }
}
procedure diffusi(quireg q)
{
    H(q);                  // Hadamard
    // Transformation
    // (superposisi)
    Not(q);                // Inversi q
    CPhase(pi,q);          // Rotate
    // if q=1111..
    !Not(q);               // undo inversi
    !H(q);                 // undo
    // Hadamard
    // Transformation
}
procedure algoritma(int bil)
{
    int sumqubit = floor(log(bil,2))+1;
    // number of
    // qubit
    int iteration = ceil(pi/8*sqrt
        (2^sumqubit));
    // number of
    // iterasi
    int rmeasurement;
    int i;
    qureg q[sumqubit];
    qureg f[1];
    print "Number of qubit
are used:",sumqubit;
```

```
print "Number of iteration
are needed:",iteration;
print "Start searching process...";
{
    reset; // clean the register
    H(q); // superposition arrangements
    for i= 1 to iterasi {
        // main loop
        print "Iteration",i;
        query(q,f,bil);
        // count C(q)
        CPhase(pi,f);
        // negation |n>
        !query(q,f,bil);
        // undo C(q)
        diffusi(q);
    }
    //oracle
    measure q,rmeasurement;
    // measurement
    if rmeasurement==bil {
        print "result of measurement:",
rmeasurement;
        print "has been equal with the
searched number...";
    }
    else {
        print "result of measurement:",
rmeasurement;
        print "has not been equal with
the search number...";
    }
    until rmeasurement==bil;
    reset; // clean the register
}
procedure start(){
    int bil;
    print;
    print "-----";
    print;
    print "SIMULATION of QUANTUM SEARCH
Using GROVER's ALGORITHM";
    print;
    input "Enter an integer
that will be searched:",bil;
    algoritma(bil);
    print;
    print "-----";
}
```

# Implementation of a new Fuzzy Based Load Balancing Algorithm for Hypercubes

Sameena Naaz

Afshar Alam

Ranjit Biswas

Department of Computer Science, Jamia Hamdard

Institute of Technology and Management

Hamdard University New Delhi, India

Gurgaon, Haryana

[snaaz@jamiahamdard.ac.in](mailto:snaaz@jamiahamdard.ac.in)

[aalam@jamiahamdard.ac.in](mailto:aalam@jamiahamdard.ac.in)

[ranjitbiswas@yahoo.com](mailto:ranjitbiswas@yahoo.com)

**Abstract—** Distributed computing systems are becoming increasingly available because of the rapid decrease in hardware cost and the advances in computer networking technologies.. It is frequently observed that in a computing environment with a number of hosts connected by a network, the hosts are often loaded differently. In typical distributed system task arrive at the different nodes in random fashion. This causes a situation of non-uniform load across the different nodes. Load imbalance is observed by existence of nodes that are highly loaded while the others are lightly loaded or even idle. Such situation is harmful to the system performance in terms of response time and resource utilization. In the work presented in this paper we have tried to analyze the effect of using fuzzy logic to deal with the problem of load balancing in hypercube model

**Keywords** Load Balancing, Fuzzy Logic, Hypercubes, Response Time

## I. INTRODUCTION

With the rapid decrease in the cost of hardware and the simultaneous increase in the computer networking technologies the increase of distributed computer systems has increased a lot. The obvious advantage of using these systems is information and resource sharing. At the same time we know that it allows parallel execution of a job on multiple processors. When the jobs are being executed in parallel on different systems a decision has to be made on to which system a newly arrived job has to be send. In a typical distributed system the jobs arrive in random fashion on different nodes. This causes a situation where some of the nodes are heavily loaded whereas others are lightly loaded. Load balancing is the technique which helps in even distribution of the jobs among the available nodes so that the throughput and the response times can be increased

Different load balancing algorithms have different complexity which depends upon the amount of communication needed to approximate the least loaded node. These algorithms can be static or dynamic in nature. Static algorithms collect no information and make probabilistic balancing decisions, while dynamic algorithms collect varying amounts of state information to make their decisions. Previous research on static and dynamic load balancing can be found in [1]-[5], [6, 7], respectively. It has been established from the previous studies that dynamic algorithms give better performance improvement as compared to static algorithms.

When we are talking about large distributed systems there is huge amount of global state uncertainty present in it. Fuzzy logic based distributed load balancing algorithms reflect the effect of uncertainty in decision making process. This approach has been discussed in [8]. The fuzzy logic approach for Distributed Object Computing Network has been studied in [9, 10]. Parallel and distributed computing environment is inherently best choice for solving/running distributed and parallel program applications. In such type of applications, a large process/task is divided and then distributed among multiple hosts for parallel computation. In [10] it has been pointed out that in a system of multiple hosts the probability of one of the hosts being idle while other host having multiple jobs queued up can be very high. In [11] the performance of a new Fuzzy Load balancing algorithm is compared with the existing algorithms.

In a distributed environment the processors are categorized according to workload in their CPU queues as heavily loaded (more tasks are waiting to be executed), lightly loaded (less tasks are waiting to be executed in CPU queue) and idle processors/hosts (having no pending work for execution). Here CPU queue length is used as an indicator of workload at a particular processor. The algorithms used for load balancing may require no information, or only information about individual jobs (static algorithm) or may make decisions based on the current load situation (dynamic algorithm).

In general, load balancing algorithm can be analyzed in a framework with four dimensions: selection policy, transfer policy, information policy, and location policy. Specifically, information and location policies have the most important roles.

- **Transfer policy:** First of all the state of the different machines is determined by calculating it's workload. A transfer policy determines whether a machine is in a suitable state to participate in a task transfer, either as a sender or a receiver. For example, a heavily loaded machine could try to start process migration when its load index exceeds a certain threshold.
- **Selection policy:** This policy determines which task should be transferred. Once the transfer policy decides that a machine is in a heavily-loaded state, the selection policy selects a task for transferring. Selection policies can be categorized into two policies: preemptive and non-preemptive. A preemptive policy selects a partially executed task.

As such, a preemptive policy should also transfer the task state which can be very large or complex. Thus, transferring operation is expensive. A non-preemptive policy selects only tasks that have not begun execution and, hence, it does not require transferring the state of task.

- **Location policy:** The objective of this policy is to find a suitable transfer partner for a machine, once the transfer policy has decided that the machine is a heavily-loaded state or lightly-loaded one. Common location policies include: random selection, dynamic selection, and state polling.
- **Information policy:** This policy determines when the information about the state of other machines should be collected, from where it has to be collected, and what information is to be collected.

## II. DISTRIBUTED SYSTEM MODEL

A simple model of a distributed system is presented here. This model consists of a decentralized decision making approach with cooperation from all the nodes. So the performance can be improved here purely by intelligent decision making and proper coordination. The various nodes of the system here are the resources and each of these resources can be in different states. A vector is used to give the state of a node which describes many characteristics of the node. The elements of this state vector are measures which imply a cost or penalty for using the resource.

The set of states of all the resources in the distributed system is known as the global system state. In distributed load balancing also the decisions are not always necessarily made using the complete global state information. In fact for each node under consideration only a subset of neighboring nodes may be needed to take a decision. Another important aspect is that a node can change state faster than the time taken to transmit state information from one state to another. Therefore there is always some amount of uncertainty in the state information used for making a decision. Hence it is necessary that the decision making process deals with these uncertainties. Fuzzy logic is one of the methods of dealing with this uncertain information and has been used in the work presented in this paper.

## III. DESCRIPTION OF ALGORITHM

The algorithm implemented in this work is basically divided into two parts:

- A. The system Model
- B. The Scheduler
  - a) Threshold Estimation
  - b) Decision Making

### A. The System Model

A distributed system is assumed to be a collection of autonomous nodes connected by a communication network. Message passing is the only form of communication among

nodes. The system model with  $d$  dimensions has  $2^d$  nodes. If we number the nodes from 0 thru  $(2^d)-1$  and look at the numbers as  $d$ -digit binary numbers, then each node will be connected to  $d$  other nodes which differ only in one bit. That is, suppose  $d=4$  for example, then 0010 (node 2) will be connected to 1010 (node 10), 0110 (node 6), 0000 (0) and 0011 (node 3). The system model is a compound module type which consists of a sender (source) and receiver (sink) module.

### B. The Scheduler

Scheduler has two functions, threshold estimation and decision making. When a scheduler is invoked, it estimates two numerical thresholds from the current states of uncertainty sources based on a fuzzy control base, and making scheduling and state update decision using fuzzy consistency model.

We have done the implementation of scheduler on MATLAB. In fuzzy logic toolbox we have taken two input parameters. The first input parameter is 'load' and the second one is 'Number of heavy Load Node' and one output i.e. 'status of load balance node'. We measure load and Number of heavy load node on a 0 to 10 scale and status of load balancing node on 0 to 20 scale.

We need to define fuzzy sets for the input parameters, 'load', and 'number of heavy load node' levels, and the output, 'status of load balancing node'. For this we define five membership functions for first input parameter i.e. 'load' and two membership functions for second input parameter i.e. 'number of heavy load node' and two membership functions for output parameter 'status of load balance node.'

### Threshold Estimation

The Threshold Estimation determines the limiting value for each membership function. Beyond this limiting value the membership function will change.

#### First Input parameter: Load (0-10)

Member Function 1:	Very lightly	(0-2)
Member Function 2:	lightly	(1-5)
Member Function 3:	moderate	(4-6)
Member Function 4:	heavy	(5-9)
Member Function 5:	very heavy	(8 -10)

#### Second Input Parameter: No. of heavy load node (0-5)

Member Function 1:	more	(0-2.5)
Member Function 2:	less	(2.5 – 5)

#### Output Parameter: Status of load balance node (0-10)

Member Function 1:	receiver	(0-5)
Member Function 2:	sender	(6-10)

In our work here we have taken the Gaussian distribution function for all the different linguistic variables for the input "load". This is shown in figure 1 below.

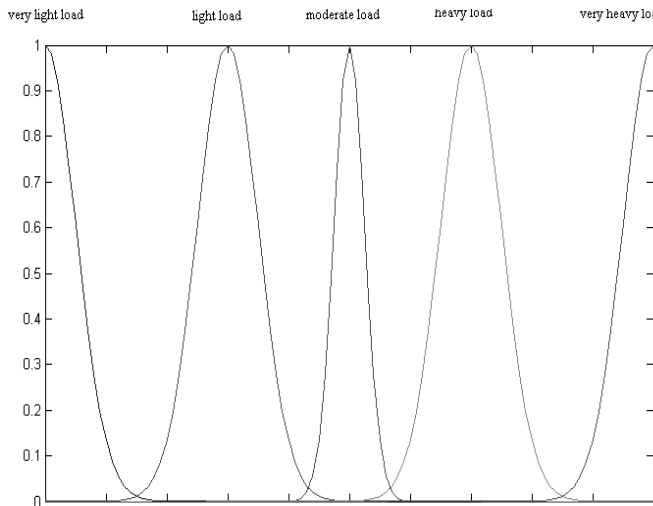


Figure 1: Input variable load of the node under consideration and it's membership function

The membership function used for the number of heavy load nodes is shown in figure2.

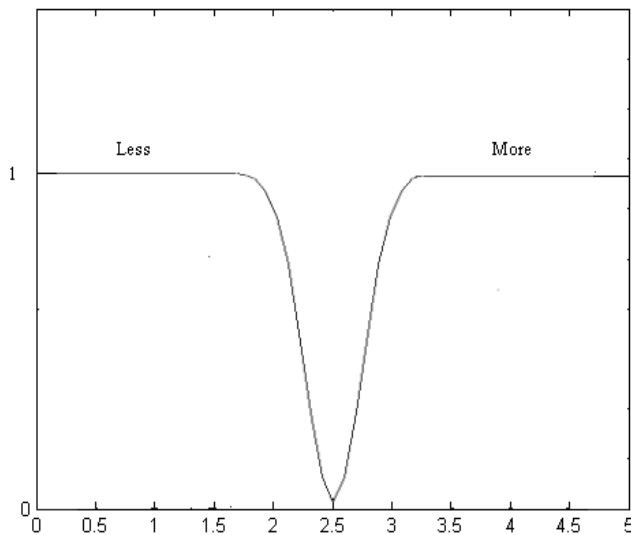
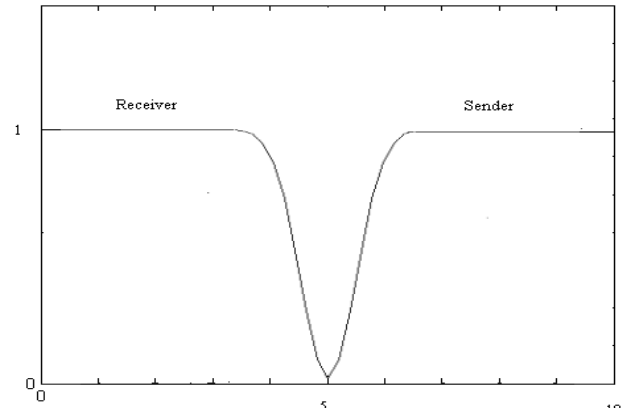


Figure 2: Input variable No. of Heavy Load Node and it's membership function.

Figure 3: Membership function for the output variable Status of Load Balance Node



The membership function for the output variable status of load balance node is shown in figure 3. From this figure we can see that there are two linguistic variables sender and receiver here and the load on a node determines it's value based upon the membership function.

#### Decision Making

The Fuzzy rules that have been used in this work are given below:

Rule [1]. If (load is very light) then (node is receiver)

Rule [2]. If (load is very heavy) then (node is sender)

Rule [3]. If (load is heavy) and (no. of heavy load nodes is less) then (node is sender)

Rule [4]. If (load is heavy) and (no. of heavy load nodes is more) then (node is receiver)

Rule[5]. If (load is light) and (no. of heavy load nodes is more) then (node is receiver)

Rule[6]. If (load is light) and (no. of heavy load nodes is less) then (node is sender)

Rule [7]. If (load is moderate) and (no. of heavy load nodes is more) then (node is receiver)

Rule [8]. If (load is moderate) and (no. of heavy load nodes is less) then (node is sender)

#### IV. INTERPRETATION OF RESULTS

A distributed system is assumed to be a collection of autonomous nodes connected by a communication network. Message passing is the only form of communication among

Node Number	Input 1	Input 2	Output
0	5	0	7.105
1	4	2	8.023
2	3	1	3.584
1	4	1	2.058
3	3	2	9.251
4	4	1	1.577
0	1	0	1.577
2	2	2	8.023
5	3	1	3.114
4	4	2	6.025
6	2	2	7.105
3	2	1	2.058
7	2	2	9.251
5	5	0	8.023
7	1	0	3.114

nodes. We have implemented system model on OPNET++ a Network Simulation Tool. The system model with d dimensions has  $2^d$  nodes. If we number the nodes from 0 through  $(2^d-1)$  and look at these as d-digit binary numbers, then each node will be connected to d other nodes which differ only in one bit. The system model is a compound module type which consists of a sender (source) and receiver (sink) and a router module.

Table 1: Output status of the nodes based upon the values of two inputs

For the hypercube model used in the work the output variable for different values of the two input variables has been shown in table 1 above. These values are used to determine whether a node works as a sender or as a receiver at any particular point of time. Once the status of the node is established random pairing between one sender and one receiver is done and the jobs are transferred. After that the response time is calculated which is depicted in table 2.

Table 2: Response time for different sender receiver pair.

#### V. CONCLUSION AND FUTURE WORK

The results of our simulation were compared with the

Sender Node	Receiver Node	Response Time with Fuzzy Approach	Response Time without Fuzzy Approach
0	1	1.99	3.45
1	3	1.99	2.96
3	2	1.99	3.66
5	7	1.99	2.99
5	1	1.00743	1.42
3	5	2.409	3.06
2	7	2.99	3.21
6	0	2.99	3.32
4	1	2.99	3.39
7	4	2.03683	3.99
4	2	2.88869	4.06
1	0	1.77632	2.01
2	0	1.7117	2.11
1	7	2.8208	4.88
7	0	3.99	4.92
5	1	1.93824	2.07

results without using fuzzy logic and from the table 2 above we can see that for all cases the response times of our algorithm is lesser than the one where fuzzy logic is not used.

In future we plan to use some more system characteristics to determine the load at a point on any node so the results are closer to the real time environment.

#### REFERENCES

- [1]. P. V. McGregor and R. R. Boorstyn, "Optimal load balancing in a computer network," in Proc. 1975 Int. Conf. on Commun., vol. 3, pp. 41.14-41.19.
- [2]. E. D. S. Silva and M. Gerla, "Load balancing in distributed systems with multiple classes and site constraints," in Proc. Performance'84, pp. 17-33.
- [3]. A. N. Tantawi and D. Towsley, "A general model for optimal static load balancing in star network configurations," in Proc. Performance'84, pp. 277-291.
- [4]. A. N. Tantawi and D. Towsley, "Optimal static load balancing in distributed computer systems," J. ACM, vol. 32, no. 2, pp. 445-465, Apr. 1985.
- [5]. J. F. Kurose and S. Singh, "A distributed algorithm for optimum static load balancing in distributed computer systems," in Proc. IEEE INFOCOM'86, pp. 458-467.
- [6]. F. Bonomi and A. Kumar, "Adaptive optimal load balancing in a heterogeneous multiserver system with a central job scheduler," IEEE Trans. Computer, vol. 39, pp. 1232-1250, Oct. 1990.
- [7]. T. C. K. Chow and J. A. Abraham, "Load balancing in distributed systems," IEEE Trans. Software Eng., vol. SE-8, pp. 401-412, July 1982.

- [8]. Chulhye Park and Jon G.Kuhl, "A Fuzzy based distributed load balancing algorithm", Proceedings of the Second International Symposium on Autonomous Decentralized Systems (ISADS'95) IEEE, 1995.
- [9]. Lap-Sun Cheung, "A Fuzzy Approach to Load Balancing in a Distributed Object Computing Network", First IEEE International Symposium on Cluster Computing and the Grid (CCGrid'01) risbane, Australia May 15-May 18, 2001.
- [10]. Yu-Kwong Kwok And Lap-Sun Cheung, "A NEW FUZZY-DECISION BASED LOAD BALANCING SYSTEM FOR DISTRIBUTED OBJECT COMPUTING", Journal Of Parallel And Distributed Computing, Volume 64 Issue 2, February 2004
- [11]. Abbas Karimi, Faraneh Zarafshan, Adznan b. Jantan, A.R. Ramli and M. Iqbal b. Saripan, "A New Fuzzy Approach for Dynamic Load Balancing Algorithm" International Journal of Computer Science and Information Security(IJCSIS).

#### AUTHORS PROFILE



Sameena Naaz received the degree of B.Sc Engg. in computers from Aligarh Muslim University, in 1998 and the M.Tech Degree in Electronics from Aligarh Muslim University, in 2000. She is currently working

as an Assistant Professor at Jamia Hamdard University in the Department of Computer Science and is also pursuing her Ph. D. Her research interests include soft computing and load balancing and scheduling in distributed systems.



# Towards a Better Assessment of the Durations in PERT Method

Nasser Eddine MOUHOU<sup>1</sup> Hocine BELOUADA<sup>2</sup> and Abdelhak BOUBETRA<sup>3</sup>

<sup>1</sup> Computer science department Sétif University ALGERIA

Tel.: +213 771 64 98 17 Fax: +213 35 66 63 10

Email: [mouhoub\\_n@yahoo.fr](mailto:mouhoub_n@yahoo.fr)

<sup>2</sup> Computer science department M'sila University ALGERIA

<sup>3</sup> Computer science department Bordj Bou Arréridj University ALGERIA

**Abstract—** During many years, two of the most popular approaches to the project management were used. It is about the critical path method (CPM) and the PERT method (Program and Evaluation Review Technique). They were based on modeling by digraphs.

CPM is unaware of the stochastic nature of the activities and brings back the model to a deterministic model. PERT holds in account this case but the estimation of the activities is despoiled with several errors.

In this paper, this technique is presented. It will be followed by an analysis, criticisms and new proposals to make corrections to this method.

**Keywords** Critical Path Method (CPM), PERT method, stochastic PERT.

## I. INTRODUCTION

To control the project is of paramount importance for organizations and is of utmost concern for the project manager confronted with a world economy where low costs and quality are the key words of the performance [1]. Since quality is not easily measurable, the main part of the research in control of project relate to the evolution of the deadlines and the costs.

Scheduling is located exactly in the planning phase of the project management. It carries out the operational follow-up: management of resources, follow-up of advance, launching of the activities. Technically, to schedule a project consists in programming in time the execution of the activities, while respecting the restrictions and so as to optimize the selected criteria of performance. It is mainly at this stage that interferes the scheduling techniques of the project presented as follows.

The fundamental techniques of scheduling, largely applied and sustained by many scheduling commercial systems, are well known such as the bar chart, the method of the potentials (MPM) and the critical path method (CPM/PERT). A variety of special techniques was also developed to solve specific circumstances or issues. With the availability of more powerful computers and software, the use of the advanced techniques becomes of greater importance for the practice. In this paper, we shall examine an extension of PERT, namely the probabilistic case. We shall then, present an analysis and criticism followed by a whole of suggestions.

## II. UNCERTAIN DURATIONSSCHEDULING

The PERT technique of three evaluations is a traditional and a well-known approach for the evaluation experts. Evaluators suggest the optimistic, pessimistic and the most probable durations. Then, a probability distribution is established to adapt the data [2].

It is however interesting to throw a critical glance on PERT method with three estimations to raise the insufficiencies which hinder the prediction validity of the tool performance in a context of interdependence of the paths of PERT network and its effectiveness in the decision-making in context of uncertainty upon the project.

In this particular case of the projects, the performance will be measured via the virtuous triangle «**cost – time – quality**» that the professionals still call «**Holy Trinity**» [3]. The performance of a project is thus measured in terms of project realization time, of the project cost and its quality.

### A. Evaluation of activity durations in an uncertain environment

The activity duration is the past time required for the achievement of this activity. According to [4], to assess the duration is probably one of the most critical devices of CPM. Several reasons urge the project manager to depart from the achievement of a rigid calendar.

For this reason, PERT suggests an assessment technique of activities based on three evaluations of a given activity duration.

[5] Explains that each of the three time evaluations presumes a static level of the resource use. These evaluations should be as good as possible because PERT outcomes depend directly on them. It is not easy to obtain accurate evaluations. It will require research, collaboration with members of planning and realization team. These three evaluations should respond to the following criteria:

1) *The optimistic time*: the period of minimum time so that the activity is achieved, i.e. time that it would take for

its achievement if all were better than anticipated (this estimation is noted  $a$ ).

2) *The most sensitive time*: the best evaluation of the period of time when the activity can be achieved, the most probable i.e. (this estimation is noted  $m$ )

3) *The pessimistic time*: the maximum period of time that would take the activity would take to be achieved, the most pessimistic i.e. (this estimation is noted  $b$ ).

These numbers ( $a$ ,  $b$  and  $m$ ) can be obtained, for example, by questioning the heads of workshops, building sites, laboratories, etc who carry out the activities.

It is acceptable to state these evaluations in days, weeks, or month as long as measurement is evenly used. When made, the evaluations of the activity time are related and should not be changed. The following reports/ratios of time must be respected:  $a \leq m \leq b$

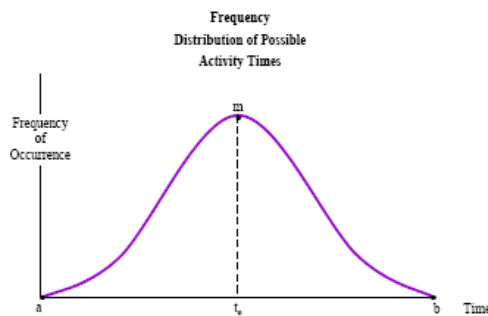


Figure 1. Time distribution of the activities drawn from [ 6].

The relative values (positions) of  $a$ ,  $m$  and  $b$  on each distribution, naturally, depends on the decision and calculations of the appreciator. When established, their relative positions on the distribution affect the value or the position of the average time of the activity noted  $t_e$  which is calculated as follows:

$$t_e = \frac{a + 4m + b}{6}$$

If one knows the distributions of the activities durations, one can calculate the average duration and the variance for each activity. However, when one has no forward statistics, one can then call upon a frequency distribution  $f(x)$ .

$$E(x) = \frac{a + b + 4m}{6}; \quad \sigma^2(x) = \left( \frac{b - a}{6} \right)^2$$

For each activity A, one determines the estimations of the quantities  $a$ ,  $b$ , then  $m$  and one deduces from the previous formulas, the estimations of its average duration  $E(x)$  and its variance  $\sigma^2(x)$ .

$$E(x) = \frac{a + 4m + b}{6}; \quad \sigma^2(x) = \left( \frac{b - a}{6} \right)^2$$

### B. Example

Let us consider the following example  $a$ ,  $b$ ,  $m$  being known for each activity, and that the following results are obtained:

ACTIVITY	$a$	$b$	$m$	DURATION	PREDECESSORS	
$\alpha$	0	0	0	0	–	0
A	7	19	13	13	$\alpha$	4
B	3	15	9	09	$\alpha$	4
C	4	40	16	18	$\alpha$	36
D	6	24	12	13	B	9
E	4	28	7	10	B	16
F	1	7	04	04	A,E	1
G	3	27	9	11	C,E	16
H	2	20	5	07	C	9
I	4	28	7	10	D	16
J	7	25	10	12	H	9
K	7	37	4	10	F	25
$\omega$	0	0	0	0	I,J,K	0
				E(x)		V(x)

Table 1. List of the prior and the durations calculated from  $a$ ,  $b$  and  $m$ .

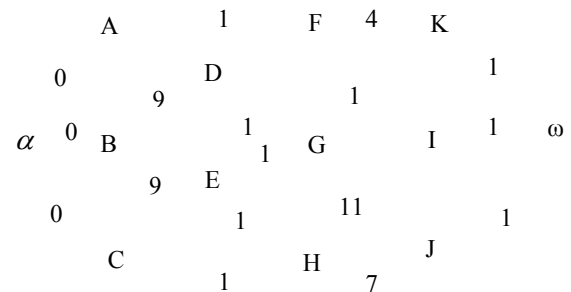


Figure 2. Activity on Nodes directed acyclic graph corresponding to table 1 with the critical path in boldface.

By using these average durations, the critical activities are  $\alpha$ , C, G, J and  $\omega$

The average duration of the project achievement is equal to the amount of the average durations of  $\alpha$ , C, G, J and  $\omega$  therefore:

$$E(x) = 0 + 18 + 11 + 12 = 41.$$

The variance is equal to the amount of the variances of  $\alpha$ , C, G, J and  $\omega$  thus:

$$\sigma^2 = 0 + 36 + 16 + 9 + 0 = 61$$

$$\text{The standard deviation } \sigma = \sqrt{\sigma^2(x)} = 7,81.$$

The reduced centered value is calculated

$$t = \frac{x - E(x)}{\sigma(x)}$$

and one can find in a table of Laplace-Gauss the probability so that the project is carried out in a given time, for example:

Project duration	Centered reduced value	Probability
35	$t = (35-41)/7,81 = -0,76$	$P(+t)=1-P(-t)$ $P(-t)=1-P(+t)$ $= 22,36 \%$
37	$t = (37-41)/7,81 = -0,51$	$P(-0,51) = 30,50 \%$
39	$t = (39-41)/7,81 = -0,25$	$P(-0,25) = 40,13 \%$
41	$t = (41-41)/7,81 = 0$	$P(0) = 50 \%$
43	$t = (43-41)/7,81 = 0,25$	$P(0,25) = 59,87 \%$
45	$t = (45-41)/7,81 = 0,51$	$P(0,51) = 69,50 \%$
47	$t = (47-41)/7,81 = 0,76$	$P(0,76) = 77,64 \%$
...		

Table 2. The probability recapitulation corresponding to each selected total duration for the project.

### III DRAWBAKS OF PERT APPROACH

One understands easily that if mistakes sully the source data of PERT to knowing **a**, **b**, **m** they accuse the righter probability to complement the project in time, taking into account the systematic skew which they make in the calculation of the expected total duration of the project and its variance [7].

In what follows is a general set of notes that seem to us fundamental and which blame the probabilistic approach of PERT.

- *Potential errors related to the inaccuracy of the estimations (a), (b), and (m)*

The estimation of the source data (**a**), (**m**) and (**b**) of PERT is in general a challenge of size for the project manager. It is frequent to note that (**a**), (**m**) and (**b**) are presented systematically in a symmetrical and a regular form, the kind of 20, 30, 40. For solving this problem, certain authors suggested to replace the absolute values (**a**) and (**b**) in percent of orders 5% and 95% respectively or 10% and 90%. From where a source of error [2]

By considering the possibility of having incorrect estimations, it seems less certain to obtain the accurate parameters of the supposed beta distribution of the activities duration.

Obviously, the estimation depends on the experiment of the responsible person. The estimations can be improbably optimistic or pessimists taking into account

the human factor or of the subjectivity which surrounds them. Besides, the relationship between the experience gained in similar projects management on the one hand and the quality of the estimations of the activities duration of an innovative project on the other hand, is perhaps not strong. For all this, it is necessary to add the risk of the durations' inflation by the manager who wants to give himself a rather broad safety margin.

- *Determinism in the determination of the critical path*

The determination of the critical path in PERT relates to the expected durations of the activities so that the stochastic element, as to the variance of the activities durations, is taken into account only at the end of the analysis in the probability calculus in order to finish the project in time [8]. As a result, the analysis of the critical path is reduced to a deterministic form such as in CPM: a determinism which hinders the probability of PERT [4].

As a consequence, the total duration of PERT project is in general smaller and never larger than the real duration of the project at time *t*. Indeed, «the delays caused by the unforeseen and unfavorable winds, are often longer than the advances due to the good luck or a zephyr collaborator» [9]. The hoped total duration of the project in PERT in general undergoes a skew towards the left of the distribution thus towards the optimistic time. It is what makes to say to Soroush [10] that the traditional approach in conventional PERT is unaware of the fact that the probability of the activities durations gives good chances to a range of paths to become potentially critical and so led to an estimation rather excessively optimistic of the probability of achieving the project in time. It is thus another source of error.

Archibald et al. [11] show that the critical path determined in PERT being not necessarily the longest, therefore not necessarily critical. In other words, the critical path determined in PERT does not include necessarily the critical activities, the latter which are supposed to be the only ones which attract the attention of the project manager if he wants to avoid any delay in the achievement of the project. It is also another source of error.

- *The interdependence of the paths or the relationship of the paths*

Two or several ways can share shared activities. They are interdependent or are thus correlated. The ways of the network are thus not always independent from the point of view of their duration, such as presupposed by PERT in the determination of the critical path.

Consequently, if a path has a very great duration, the other paths with which it shares activities could also have a long duration [5]

- *Potential errors related to the assumption of Beta distribution*

One of the fundamental assumptions of PERT is Beta distribution of the activities duration (fig. 3.).

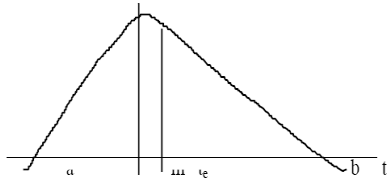


Figure 3. Curve of the distribution Beta drawn from [7].

Archibald et al. [11] recall that since the accurate form of the distribution of the activities durations remains probably unknown and considering the features alleged distribution of an activity to presume a priori that the distribution of an activity duration following a Beta law is potentially an error source.

Indeed, other distributions of completely different shapes also possess the properties of the beta distribution, while having a very different average and variance. Consider for example the distributions shown in figure 4. The four distributions have the same mode  $m$  and a similar extent  $[0, 1]$ . The distribution  $D_1$  is a beta distribution.  $D_2$  is a nearly uniform distribution,  $D_3$  is a function almost delta and  $D_4$  is almost zero between 0 and  $m$  and almost uniform on the interval  $[m, 1]$  where we assume  $0 \leq m \leq 1/2$  (if  $m$  is in the other half, the results remain valid, by symmetry).

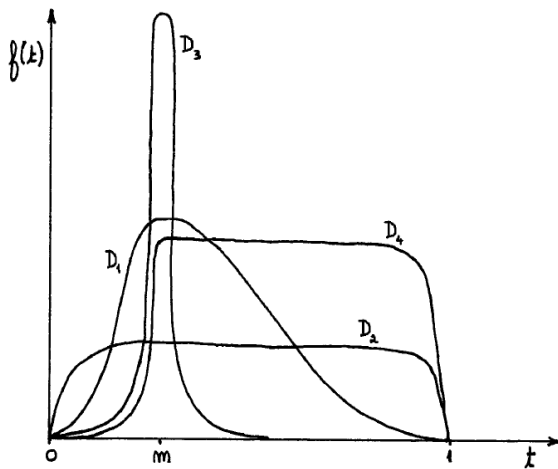


Figure 4. Example of possible distributions.

The beta distribution  $D_1$  has an average of  $(4m + 1) / 6$  (since  $a = 0$  and  $b = 1$ ) and a standard deviation of  $1 / 6$  of interval is considered. For an almost uniform distribution  $D_2$ , we have an average close to half and a standard deviation that is close to  $1 / 12$ . The distribution near Delta  $D_3$  has an average close to the mode  $m$  and variance close to zero. For distribution  $D_4$ , the average may be very close to  $(m + 1) / 2$ . Although the previous three possible distributions are unlikely to be met (these extreme cases that we want to compare to the chosen beta distribution), they allow us to obtain a bound on the possible error caused by the use of inappropriate distribution. We will therefore calculate the maximum possible difference between the average for the chosen

beta distribution and other distributions. Thus, the maximum absolute error on the average is 8 [5].

$$\max \left[ \left| \frac{(4m+1)}{6} - \frac{1}{2} \right|, \left| \frac{(4m+1)}{6} - m \right|, \left| \frac{(4m+1)}{6} - \frac{(m+1)}{2} \right| \right] = \frac{(1-\frac{m}{2})}{3}$$

Based on our distributions  $D_2$  and  $D_3$  for which we have an estimate of the standard deviation, we also find that the maximum absolute error on the standard deviation is

$$\max \left[ \left| \sqrt{1/12} - 1/6 \right|, \left| 0 - 1/6 \right| \right] = 1/6.$$

The error on the average is based on the mode. If the  $m$  is very central (close to half), the error on the average is around 25%, however if it is close to bound distribution, the error can be 33%. The standard deviation error does not depend on the mode and is about 17%. Note that we have discussed the absolute error; the difference can be either negative or positive

The average and the variance of the Beta distribution are not enough by themselves to characterize a single and identical Beta distribution of the activities duration, the requirement of the asymmetry for excluding the standard model where the average and the variance determine a single distribution.

However for practical reasons, it proved to be inevitable to characterize the duration of an activity by its average and its variance only. This is another source of error [12]

Third source of error: distributions such as the triangular distribution which satisfy very much and sometimes better the features referred to above, having averages and variances quite different from those of the Beta law.

However, the assumption of the Beta distribution does not carry too much the consequence as it fulfills the requirement of the asymmetry of the density function and explains the expression of the expected duration according to the mode ( $m$ ) and of the extreme values ( $a$ ) and ( $b$ ). Anyway, the errors related to the Beta distribution are considered to be negligible compared to the following errors:

- Potential errors related to  $\sigma$  and to  $t_e$

By taking the minimal duration of the activity if all occurred ( $a$ ) exceptionally well, the duration of the activity if one encountered difficulties under development ( $b$ ) and the most probable duration, most awaited by the manager of project ( $m$ ), PERT determines the hoped duration ( $t_e$ ) of each activity and its variance. It is indeed more interesting to trust the value of the hoped time of the activities than to stick to the extreme values of the distribution of the durations of the activities, the generalizing and compensation average being.

In probability theory, the hypothesis follows  $\sigma = (a-b)/6$ . Indeed, the probability that a normal random variable  $X$  takes value  $x$  belongs to a defined interval by  $\mu \pm \lambda\sigma$ , which is given in Table 1 for  $\lambda = 1, 2, 3$ . [13]

To calculate the total expected duration of the project and its variance, the hypothesis of independence of the activities durations has been chosen for reasons of

convenience and application of the recommended central limit theorem when the range of activities on the critical path is large (30 and over).

$\lambda$	Interval $(\mu \pm \lambda\sigma)$	Probability (%)
1	$\mu - \sigma < x < \mu + \sigma$	68.3
2	$\mu - 2\sigma < x < \mu + 2\sigma$	95.4
3	$\mu - 3\sigma < x < \mu + 3\sigma$	99.7
4	$x < \mu - 3\sigma$ ou $x > \mu + 3\sigma$	0.3

Table 3. Probability  $P(\mu - \lambda\sigma \leq x \leq \mu + \lambda\sigma)$

Based on these assumptions, we obtain the respective estimates of expected time ( $t_e$ ) and a standard derivation ( $\sigma$ ) of the duration of each activity:

$$t_e = (a + 4m + b) / 6 \quad \text{and} \quad \sigma = (b - a) / 6$$

So that, assuming the length of a sequence of activities is equal to the sum of the durations of each of the activities in the sequence. Hence the total of independent random variables of beta distribution and applying the central limit theorem, PERT considers that the total project duration is the total of the expected time of critical activities and the variance of the total project duration is the sum of the variances of the durations of critical activities [13].

While considering a priori that the distribution of the duration of an activity follows a law beta and knowing that the expression of the function of density of this law is well known, one exposes oneself to two sources of errors: *First source of error*, the variation - type of the duration of an activity was supposed to be equal to  $(b - a)/6$  as previously explained. It is certainly not the true value of the variation - type of a Beta distribution!

*Second source of error* the linear approximation  $t_e = (a + b + 4m)/6$  made for the true value of the hoped time which is rather a root of a cubic equation. From where the incompatibility of the formulas of hoped time and the standard deviation in practice [2].

- *Errors related to the assumption of independence of the activities durations.*

The assumption of independence of the durations of the critical activities is less expected. The toughness of the climate for example, can affect in the same way several activities and lead the person liable to accelerate the project. The net effect of the inflationary and deflationary factors over the total duration of the project would be neglected if activities durations were supposed to be independent [14]

#### IV. SOME SUGGESTIONS

##### 1. A technique PERT with only two estimations.

Being given the serious difficulties experienced by the project managers to obtain the estimation of the most probable time until put into practice, «the most

probable time is generally the most improbable time », would not it be indicated to consider a technique of PERT with two estimations: the optimistic duration ( $a$ ) and the pessimistic one ( $b$ )? If we consider a Beta distribution of the activities duration, Golenko- Ginzburg [14] observes that the estimation of the most probable duration ( $m$ ) on the whole of projects network is «practically useless» and that their value ( $m$ ) is often close to  $(2a + b)/3$ , in addition, there is no significant statistical difference in using the values of ( $m$ ) rather than the calculated values. Thus the following formulas of expected durations and variances are deduced:

$$t_e = 0.2(3a + 2b) \quad \text{et} \quad \sigma^2 = 0.04(b - a)^2$$

PERT with two estimations has the advantage of simplifying the analysis without affecting the project parameters and offers the opportunity of a combined approach PERT/CPM for the management of the project deadlines/cost, whatever is its environment.

##### 2. PERT/CPM combined approach and analysis of the compromise between resource allocation and the risk of project delay.

PERT and CPM are two convergent and synergistic techniques. Why don't we try to combine them to take advantage of the forces and to correct their weaknesses? MacLeod and Petersen [16] are bent over to study the question.

We notice that PERT and CPM rest on the whole on only one estimation of time which is the expected time ( $t_e$ ) in PERT and normal time ( $n$ ) in CPM and that PERT and CPM have also a lower limit of duration for the activities of the project, hence the minimal time of achievement of the activity ( $c$ ) in CPM and the optimistic time ( $a$ ) in PERT, one can validly accept with MacLeod and Petersen [16], the assumption  $t_e = n$  et  $a = c$ . In a word, the calculated expected times in PERT are for PERT, the standard times are for CPM and the activity minimal duration is for CPM what the optimistic duration is for PERT.

With such an assumption, it is henceforth possible to calculate the effects of additional resource allocation on the probability of finishing the project in time. Hence PERT/CPM combined approach makes it possible to profit from the analysis of the compromise deadlines/cost that only CPM allows and to evaluate the opportunity to achieve the project in time (that is not possible with PERT), as a whole when using a system based on the compromise between resource allocation and the risk of project delay.

##### 3. Taking into account the assumption of the dependence of the paths durations in the determination of the project total duration and its variance.

The assumption of independence of the paths durations is error-prone in the evaluation of the risk of the project delay. It is what justifies the opportunity in taking into account of the opposite assumption of dependence of the paths durations. Obviously by considering the most

plausible assumption of correlation of the paths durations of the network, one improves the chances to obtain a true probability to finish the project in time. Taking into account only the interdependence or the correlation of the paths durations, is to agree to face a tiresome and complex calculation of the opportunity to finish the project within the time  $t$  knowing the multi-activities nature and multi-paths of the network.

Several researchers studied the issue and suggested resorting if the project scope authorizes it, to the simulation of Monte Carlo, or an approximation by the heuristic ones or limiting the probability of finishing the project on a date  $t$  given in an interval. This last method (the algorithm of Probabilistic Network Evaluation Technique) is stochastic and the lower and higher limits are respectively the probabilities for finishing the project on the assumptions of independence and statistical paths dependence [17]

**4. The construction of a reliable interval around the PMB (Performance Measurement Baseline) under the conditions of a Beta distribution of the activities duration, interdependence of the durations of the paths, of PERT with two estimated and central limit theorem.**

The *PMB* is the planned cost (**P**), set aside of the provisions and the overheads. As such, the *PMB* is directly related to the project acquired or gained value which is (**V**). From a practical point of view, the value is gained only if the parts of the work were delivered [2]

## V. CONCLUSION

The acquired value analysis is based on a core planning PERT / CPM. In this article, we stressed the crucial mistakes that undermine the chances for obtaining a good measure, whether the errors associated with data PERT or those related to the particular network configuration of the project.

The proposed approach could help reduce such errors. From a theoretical viewpoint, it becomes possible to appreciate the significance of the variances of time and probability of realization of the value acquired at any time.

In this effort to provide the project manager of a strong tool for monitoring projects, we should not lose sight of the issue of complexity. Accept, for example, to take into account the hypothesis of interdependence of network paths, it is undoubtedly to face more complexity in the acquired value method. Also, should we admit that the trade-off analysis of cost/time is not adequate for the activities for which the addition of extra resources does not help to accelerate them? This is a first restriction of the approach.

The linearity assumption of the project activities costs is also restrictive in that it ignores the case of activities for which expenses are incurred only at the end of achievement. Thus, appears a second restriction of the approach.

But the most important, is perhaps the interest of such an approach in practice. We may indeed wonder what the

marginal benefit of such an approach is, given its complexity. Is the game worthy for such an approach? This is an important concern to be taken into account in further research. At the very least, a future work will provide a rigorous procedure, to illustrate it using digital processing and to demonstrate a real contribution to the project manager.

## REFERENCES

- [1] J. COUILLARD, D. GODCHARLES & J-P. PAQUIN, La qualité acquise: pour un meilleur contrôle des projets, *Revue Internationale en gestion et management de projet*, vol 3, n° 1, 46-59, (1998).
- [2] C. PESCIO, Realistic and Useful: Toward Better Estimates, Draft version 1.2 - April, (2007).  
Online reference:  
<http://www.eptacom.net/betterestimate/TheMathBehind.pdf>
- [3] J. P. KINDINGER, & J. L. Darby, Risk Factor Analysis – A New Qualitative Risk Management Tool, *Proceedings of the Project Management Institute Annual Seminars & Symposium*, Houston, Texas, USA, (2000).
- [4] G. RULE, Bees and the Art of Estimating, *IEEE Software*, November/December (2000), p. 23.
- [5] D. DUBOIS, H. FARGIER, & V. GALVAGNAN, On latest starting times and floats in activity networks with ill-known durations. *European Journal of Operational Research* 147, (2003).
- [6] A. KIMMS, Mathematical Programming and Financial Objectives for Scheduling Projects, *Kluwer Academic Publishers*, Boston, (2001).
- [7] L. A. IKA, L'analyse de la valeur acquise en contexte d'interdépendance des chemins : une analyse à explorer, Université du Québec en Outaouais, *ASAC*, (2004).
- [8] P.B. ROBINSON, The Performance Measurement Baseline- A statistical view, *Project Management Journal*, vol. 28, no 2, 47 –52, (1997).
- [9] Y. NORBERT, R. OUELLET, & R. PARENT, *La recherche opérationnelle*, 2<sup>ème</sup> éd., Gaëtan Morin Éditeur, (1999).
- [10] H.M. SOROUGH, The most critical path in a PERT network: A heuristic approach, *European Journal of Operational Research*, vol. 78, no1, 93-105, (1994).
- [11] R.D. ARCHIBALD, & R.L. VILLORIA, *Network-based management systems (PERT/CPM)*, New York: John Wiley & Sons, (1967).
- [12] M. JORGENSEN, Realism in Assessment of Effort Estimation Uncertainty: It Matters How You Ask, *IEEE Transactions on Software Engineering*, April (2004).  
Online reference (draft)  
[www.simula.no/departments/engineering/artifacts/framingsubmit.pdf](http://www.simula.no/departments/engineering/artifacts/framingsubmit.pdf)
- [13] P. Aquino, A PERT approach to cost risk analysis, *AACE 1992 Transactions*. F.4.1- F.4.7, (1992)
- [14] M. JORGENSEN, K. H. TEIGEN & K. MOLOKKEN-OSTVOLD, Better sure than safe Overconfidence in judgment based software development effort prediction intervals, *Journal of Systems and Software*, Vol. 70 Nr. 1-2, (2004).



Online reference:

<http://www.simula.no/departments/engineering/.artifacts/pireviseddesember\ copy.pdf>

[15] D. GOLENKO-GINZBURG, On the distribution of activity time in PERT, *Journal of Operational Research Society*, XXXIX, vol. 39, n08, 767–771, (1988).

[16] K.R. MACLEOD & P.F. PETERSEN, Estimating the tradeoff between resource allocation and probability of on

– time completion in project management, *Project Management Journal*, vol 27, n° 1, 26– 33.

[17] J. ABDELNOUR, A. H-S ANG, & A.A. CHAKER, Analysis of activity networks under uncertainty, *Journal of the Engineering Mechanics Division, ASCE*, 101, 373-387, (1975).

# Network Anomaly Detection and Visualization using Combined PCA and Adaptive Filtering

Altyeb Altaher, Sureswaran Ramadass  
NAv6 Center of Excellence, Universiti Sains Malaysia  
USM, 11800 Penang, Malaysia  
altyeb@nav6.usm.my,sures@nav6.usm.my,

Noureldien Abdelrahman , Ahmed Khalid  
Faculty of Computer Sciences and IT, University of  
Sciences and Technology, Sudan  
noureldien@hotmail.com ,asalih2@hotmail.com

## Abstract

In recent years network anomaly detection has become an important area for both commercial interests as well as academic research. This paper provides a Combined Principal Component Analysis (PCA) and Filtering Technique for efficient and effective detection and identification of network anomalies. The proposed technique consists of two stages to detect anomalies with high accuracy. First, we apply the Principal Components Analysis to transform the data to a new coordinate system such that the projection on the coordinate contains the greatest variance. Second, we filter traffic to separate between the normal and anomalous traffic using adaptive threshold. Our analysis results from network-wide traffic datasets show that our proposed provides high detection rate, with the added advantage of lower complexity

**Keywords-** Network anomaly detection, principal component analysis , network anomaly visualization, adaptive network traffic filter.

## I. INTRODUCTION

Detecting unexpected changes in traffic patterns is a topic which has recently received much attention from the network measurement community. Network traffic is often seen to exhibit sudden deviations from normal behavior. Some of these deviations are caused by malicious network attacks such as Denial-Of-Service or viruses, whereas others are the result of equipment failures and accidental outages [1]. Heady et al.[8] defined an intrusion as “any set of actions that attempt to compromise the integrity , confidentiality or availability of information resources”. The identification of such a set of malicious actions is called intrusion detection problem that has received great interest from researchers. Several schemes proposed in the literature are derived from classical time series forecasting and outlier analysis methods and applied to the detection of anomalies or faults in networks [9, 10, 11].

Principal Component Analysis [3] (PCA) is a good statistical-analysis technique for detecting network traffic anomalies. PCA is used to separate the high-dimensional space occupied by a set of network traffic measurements into two disjoint

subspaces corresponding to normal and anomalous network conditions. The main advantage of this approach is that it

exploits correlations across links to detect network- wide anomalies. Recent papers in networking literature have applied PCA to the problem of traffic anomaly detection with promising initial results [4, 2, 5, 1].

The proposed approach consists of two stages to detect anomalies with high accuracy. First, we apply the Principal Components Analysis to transform the data to a new coordinate system such that the projection on the coordinate contains the greatest variance. Second, we filter traffic to separate between the normal and anomalous traffic using adaptive threshold.

## II . Combined PCA and Adaptive Filtering Approach

This section presents the proposed approach. In Section A, we describe the PCA based intrusion detection that is utilized for detecting the anomaly traffic. In section B we describe the Combined PCA and Adaptive Filtering Approach.

### A. Principal Component Analysis

Principal Component Analysis (PCA, also called Karhunen-Loeve transform) is one of the most widely used dimensionality reduction techniques for data analysis and compression. It is based on transforming a relatively large number of variables into a smaller number of uncorrelated variables by finding a few orthogonal linear combinations of the original variables with the largest variance. The first principal component of the transformation is the linear combination of the original variables with the largest variance; the second principal component is the linear combination of the original variables with the second largest variance and orthogonal to the first principal component and so on. In many data sets, the first several principal components contribute most of the variance in the original data set, so that the rest can be disregarded with minimal loss of the variance for dimension reduction of the data [6, 7]. The transformation works as follows.

Given a set of observations  $x_1, x_2, \dots, x_n$ , where each observation is represented by a vector of length  $m$ , the data set is thus represented by a window  $X_{n \times m}$

$$X_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = [x_1, x_2, \dots, x_n] \quad (1)$$

The average observation is defined as

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

The deviation from the average is defined as

$$\Phi_i = x_i - \mu \quad (3)$$

The sample covariance matrix of the data set is defined as

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T = \frac{1}{n} A A^T \quad (4)$$

Where  $A = [\Phi_1, \Phi_2, \dots, \Phi_n]$

To apply PCA to reduce high dimensional data, eigenvalues and corresponding eigenvectors of the sample covariance matrix  $C$  are computed. We choose the  $k$  eigenvectors having the largest eigenvalues. Often there will be just a few large eigenvalues, and this implies that  $k$  is the inherent dimensionality of the subspace governing the "signal" while the remaining  $(m - k)$  dimensions generally contain noise [7].

We form a  $m \times k$  matrix  $U$  whose columns consist of the  $k$  eigenvectors. The representation of the data by principal components consists of projecting the data onto the  $k$ -dimensional subspace according to the following rules [7]

$$y_i = U^T (x_i - \mu) = U^T \Phi_i \quad (6)$$

### B. The Proposed Detection Approach

Principal component analysis has been applied to the intrusion detection as a data reduction technique not as an anomaly identifier. In this paper we combine the PCA with adaptive filter to identify anomalies in network traffic.

Based on statistical analysis, we assume that the used data set has a normal distribution, we propose suitable analysis

and detection techniques to detect anomalies with high confidence while reducing the false acceptances

The proposed Combined PCA and Adaptive Filtering Approach consist of the following steps:

1) Fix the window size equal to  $N$  and (In our simulations we used  $N=41$ )

2) Apply the PCA in network traffic window to identify patterns in network traffic, and express the network traffic in such a way as to highlight their similarities and differences.

3) Calculate the mean and the standard deviation of network traffic window

4) if the network traffic in the window exceeds the threshold  $\Omega$  it considered as anomaly.

The threshold  $\Omega$  defined as follow

$$\Omega = \mu + c\sigma \quad (7)$$

Where  $c = 2.25$

## III. EXPERIMENTS

### A. Data

We used the Abilene dataset, this dataset was collected from 11 core routers in the Abilene backbone network for a week (Dec. 15 to Dec. 21, 2003). It comprises two multivariate time series, one being the number of packets and the other the number of individual IP flows in each of the Abilene backbone flows (the traffic entering at one core router and exiting at another), binned at five minute intervals. Both datasets,  $X(1)$  and  $X(2)$ , are of dimension  $F \times T$ , where  $T = 2016$  is the number of time steps and  $F = 121$  is the number of backbone flows[2].

### B. Anomaly Detection using the combined PCA and Adaptive Filtering

To gain a clearer understanding of the nature of the Abilene data set, we examine the Histogram of the Abilene data set as in Fig 1. the shape of histogram indicates that data is normally distributed, as a normal distribution is characterized by its bell shape. The curve of histogram is concentrated in the center and decreases on either side, this means that the data set has less of a tendency to produce unusually extreme values. Fig 2 is a plot of Abilene data set.

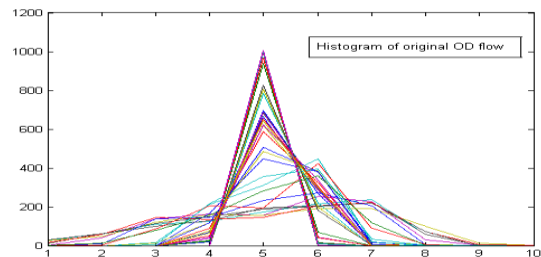


Fig 1: Histogram of original OD flow from Abilene data

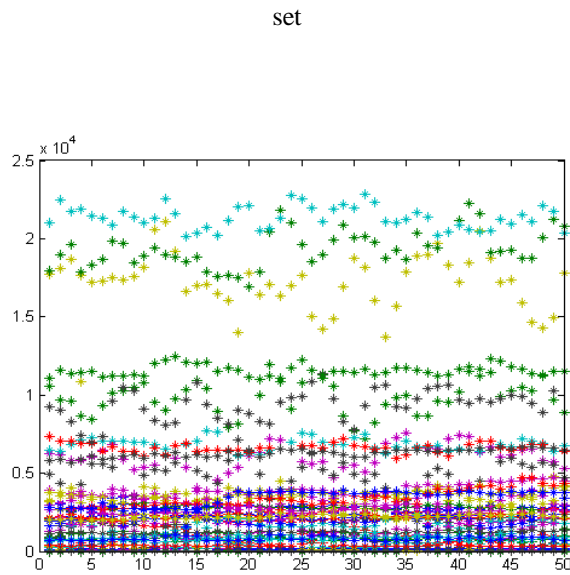


Fig 2: Plot of original OD flow from Abilene data set

We implement our proposed detection method using MATLAB, which is a high-level technical computing language and interactive environment for algorithm development, data visualization and analysis.

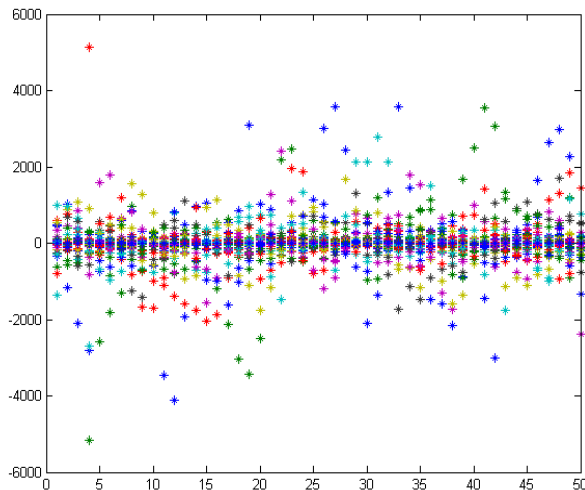


Fig 3: Plot of data after applying our proposed method.

Figure 3 is a plot of Abilene data after applying our proposed method, the normal traffic is centered in the middle, while anomalies deviates from the behavior or normal traffic, it tends to scatter far from the center.

To evaluate our algorithm, we examined its performance on network-wide traffic datasets analyzed by Lakhina et al. in [2], with well known and identified anomalies, thus we have “ground truth” anomaly annotations against which to compare the output of our Combined PCA and Adaptive Filtering Approach. We found that our Combined PCA and Adaptive Filtering method provides high detection rate, with the added advantage of lower complexity. The experimental results show that our Combined PCA and Adaptive Filtering method detects 85% of anomalies in the Abilene data set.

#### REFERENCES

- [1] A. Lakhina, M. Crovella, and C. Diot, “Mining Anomalies Using Traffic Feature Distributions,” in Proc. SIGCOMM, Philadelphia, PA, Aug. 2005.
- [2] Lakhina, A., Crovella, M., and Diot, C. Diagnosing network-wide traffic anomalies. In ACM SIGCOMM (Portland, Oregon, USA, 2004), pp. 219–230.
- [3] Hotelling, H. Analysis of a complex of statistical variables into principal components. J. Educ. Psy. (1933), 417–441.
- [4] Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E. D., and Taft, N. Structural analysis of network traffic flows. In ACM SIGMETRICS (New York, NY, USA, 2004), pp. 61–72.
- [5] Lakhina, A., Crovella, M., and Diot, C. Characterization of network wide anomalies in traffic flows. In ACM Internet Measurement Conference (Taormina, Sicily, Italy, 2004), pp. 201–206.
- [6] I.T. Jolliffe, “Principal Component Analysis”, 2nd Ed., Springer-Verlag, NY, 2002.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork, “Pattern Classification”, China Machine Press, Beijing, 2nd edition, 2004.
- [8] R. Heady, G. Luger, A. Maccabe “The architecture of network level intrusion detection system”, Technical report, Computer Science Department, University of New Mexico, August 1990.
- [9] F. Feather, D. Siewiorek, and R. Maxion. Fault detection in an ethernet network using anomaly signature matching. In Proceedings of ACM SIGCOMM, 1993.
- [10] I. Katzela and M. Schwartz. Schemes for fault identification in communication networks. IEEE/ACM Transactions on Networking, 3(6), Dec. 1995.
- [11] M. Thottan and C. Ji. Anomaly detection in IP networks. IEEE Transactions in Signal Processing, 51(8), Aug. 2003.

## A Two Dimensional Approach to Back and Forth Software Process Model

Sameera Amjad, Wasi Haider Butt, Farooque Azam

Department of Computer Engineering

College of Electrical & Mechanical Engineering

National University of Science & Technology (NUST)

Rawalpindi, Islamabad

{sameera,wasi}@ce.ceme.edu.pk, farooq@ceme.nust.edu.pk

**Abstract**—Many software development process models have been documented but none of them gives a detailed methodology for change management. This article proposes a novel software development process model which realizes the inherent nature of requirement changes and provides a methodology to accommodate these changes. A detailed literature survey was conducted to explain the difference of the proposed model with existing software development approaches. The proposed novel model namely, the Back and Forth software process model uses two methods to present the development methodology.

**Keywords**- *Software Development Life Cycle; Software Process Models*

### I. INTRODUCTION

A software development process is a well structured set of activities and associated results that are followed for the successful development of a software product. "Software Process Model is an abstract representation of a software process"[1]. The workflow model shows the sequence of activities in the process along with their inputs, outputs and dependencies. This workflow is a guideline for successful development and deployment of the software project. The development lifecycle of software comprises of four major stages namely requirement elicitation, designing, coding and testing. A process model is chosen according to the nature of the application, tools and techniques available and expertise of development execution [2].

### II. EXISTING SOFTWARE PROCESS MODELS

Listed below are some traditional and most commonly used software process models:

#### A. Build and Fix Model:

The evolution of software process models began its journey from the most primeval process model namely the "Build and Fix" model [3,15]. This model constitutes two basic

steps:

1)Writing the Code

2)Fixing Problems in the code.

To give a clearer picture of its flaws, some important points are listed below:

- Does not follow any proven method. Its working included, first coding then moving towards other stages. Due to which the resulting product had a structure which often did not meet the requirements of the user consequently ending up in either project termination or redevelopment which was highly expensive [3].
- Due to a number of fixes, the resulting code had a poor structure also these fixes were highly expensive when addressed late in the development process. This was due to the absence of a detailed design phase before the coding phase [4].
- Not suitable for environment where changes are dynamic in nature [5].
- The model handled lightly the testing phase after coding the basic requirements. This eventually led slipping of numerous undiscovered errors in the code; weakening its output [6].is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

#### B. Waterfall Process Model:

The Classical Life Cycle or the Waterfall Process Model [5] was the first process model to present a sequential framework, describing basic stages that are mandatory for a successful software development model. It formed the basis for most software development standards and consists of the following phases: Requirements elicitation, Designing, Implementation and Testing.

The model restricted software engineers to follow a sequential order moving from one stage to the next only after the completion of the former. Listed below are some flaws:

- Rigid design and inflexible procedure [5].
- Restricting back & forth movement from a later stage to a former one. When new requirements surface accommodating those with existing ones become difficult due to restrictions in looping back to prior stages.
- Waterfall Model faced “*inflexible point solutions*” which meant that even small amendments in the design were difficult to incorporate later in design phase.
- As the requirements were frozen before moving to the design phase, using the incomplete set of requirement, a complete design was worked on. Such an approach worked normally well for a small project requiring average amendments. In case of a large project, completing a phase and then moving back to reconstruct the same phase, incurred a large overhead. [5].
- Once a phase is done, it is not repeated again that is movement in the waterfall goes one to the next and the vice versa is not supported. Deadlines are difficult to meet in case of large projects. [6]

#### C. Prototype Model:

In Prototype Model [7], the user is given a “look and feel” of the system using a prototype. The prototype for the system to be developed is built, tested and reworked as necessary. Prototype process model is suitable for dynamic environment where requirements change rapidly. The process begins with gathering main functional requirements; this is followed by a quick design leading to the development of a prototype. The prototype is then evaluated by users and customers. Developers rework on the prototype until the customer and users are satisfied. The prototype can face the following limitations:

- The main limitation of this model includes lack of information about the exact number of iterations and the time period required to upgrade the prototype in order to bring it up to the satisfaction of the user and the customer [7].
- Developers are in such a rush that they hardly consider all the functionalities of the prototype. In order to release the product as soon as possible, the prototype with some additions is released on or before the target release date. This happens due to lack of user analysis

activities; the end product contains features the user is hardly aware how to use. [6]

- Often the developers make implementation compromises in order to make the prototype work quickly, which will lead to the use of inappropriate operating system or programming language [8].
- The premature prototypes lack key consideration like security, fault tolerance, distributed processing and other such key issues [3].

#### D. Incremental Development Model:

First In incremental development process [1], customers identify, in outlined the services to be provided by the system. They identify which of the services are most important and which are least important to them. A number of delivery increments are then defined which each increment providing a subset of functional requirements. The highest priority functional requirements are delivered first. The disadvantages of the model are:

- It is difficult to map requirements directly to different increments. Include excessive user involvement. Poorly defined scope as scope of product may vary increment to increment.[7]
- An overhead in the model is rapid context switching between various activities. Each iteration is followed by an evaluation ensuring that user requirements have been met [9]. This evaluation after each iteration is time consuming.

#### E. Spiral Model:

In Spiral model [10] instead of presenting a sequence of activities with some backtracking from one activity to the other, the process model followed a spiral organization of activities. It combines characteristics of both prototype and waterfall process model. The model is divided into some task regions, which are as follows: Customer Communication, Planning, Risk Analysis, and Engineering, Construction and release and Customer evaluation. The distinctive feature of this model is that each stage is controlled by a specific risk management criteria ensuring decision making using critical factors. The following disadvantages are identified in this model:

- A number of risks, constraints, alternatives, models etc. need to be analyzed but never are these risks or objectives listed and no specific risk analysis technique is mentioned. If risk analysis is poor the end product will surely suffer.



- Another difficulty of the spiral model is adjustment of contract deadlines using the spiral model.
- Risk analysis expertise is vital. For large projects expert software developers can produce efficient software products but in case of a complex large project absence of specific risk analysis techniques and presence of varying expertise can create a chaos [10].

#### F. Win Win Spiral Model:

In Win Win Spiral Model [2], the developer and customer end up in negotiating various requirements based on functionality, performance, cost, time, etc. The best negotiation strives for a “win-win” result. The detailed risk analysis imposing many different constraints, objectives & alternatives consume a lot of time. But never are these risks specifically mentioned and vary project to project [6].

#### G. Rapid Application Development Model:

The RAD model [11] is an adaptation of the classical model for achieving rapid development using component based construction. If requirements are well understood with a well constrained project scope, the RAD process enables delivery of the fully function system. The model is considered to be incremental development model and that have emphasis on short development cycle. Rapid Application Development has following drawbacks:

- Reduced features occur due to time boxing, where features are delayed to later versions in order to deliver basic functionality within abbreviated time.
- Reduced scalability occurs because a RAD developed application starts as a prototype and evolves into a finished application using existing component and their integration.
- RAD, for large projects, requires a sufficient number of human resources also requiring existence of components for reuse. Also RAD is not suitable for all types of application development [11].
- High technical risks discourage RAD use. This is because use of new technology in a software is difficult in a changing global software market [11].

#### H. Rational Unified Process:

The RUP [1] provides dynamic, static and practice perspectives of a product. The RUP provides each team member with the guidelines, templates and tool mentors necessary for the entire team to take full advantage of the best practices. The software lifecycle is broken into cycles,

each cycle working on a new generation of the product. This phased model identifies four discrete phases:

- Inception phase
- Elaboration phase
- Construction phase
- Transition phase

The identified drawbacks of the process are:

- Each phase has a milestone which needs to be satisfied for the next particular phase to start.
- If the respective milestone of the particular phase is not satisfied the entire project might get cancelled or re-engineered before proceeding further.
- The satisfaction criteria of a particular milestone has its own constraints and are not listed specifically [12].

#### I. The V-Model:

The V-Model [13] is an extension to the Waterfall Model in that it does not follow a sequential mode of execution rather it bends upward after the coding phase to form V shape.

It has the following drawbacks;

- It addresses software development within a project rather than a whole organization.
- The V-Model is not complete as it argues to be, as it covers all activities at too abstracts level.

### III. BACK & FORTH SOFTWARE PROCESS MODEL

The proposed BNF is a software development process model which can be manifested as an entire software development framework to build a centrally controlled system.

BNF software process model proceeds linearly if no new change surfaces. The moment such changes are encountered, BNF iterates back to the first phase of the development model and then proceeds linearly henceforth accommodating the new changes. Incremental iteration between immediate successive phases is also a yet another feature of the BNF software process model. These iterations are observed only if required. Successive phases work closely together filling minute details for development. BNF is based on change management and deliverable oriented approach to development.

- **BNF Principles**

BNF principles include the following:

- i. **Controlled iterations between successive phases of BNF software development process:** This principle helps in controlling the number of

iterations required to accommodate new changes depending on the deliverable deadline estimate.

- ii. **Focus on change management early to maximize stakeholder value:** New requirements may surface during later phases of software development which should be accommodated as soon as they appear to maximize user satisfaction.
- iii. **Focus on modular deliverables within deadlines:** This principle focuses on incremental development of ERP system in modules.
- iv. **Encourage stakeholder participation:** This principle promotes practices that allow project participants and stakeholders to develop a solution that maximizes stakeholder benefits, and is compliant with constraints placed on the project [14].

#### A. How BNF Process is organized

BNF software process model can be organized into two correlated dimensions: **method content** and **process content**. The method content is where only the method elements are defined (namely roles, tasks or sub phases, and artifacts). The process content explicitly specifies where the elements of method content are applied in a predefined and chronological manner.

##### 1) Method Content:

Method Content contains all the resources that will participate, get consumed or produced in the BNF software process model for the development of software. Given below are three resources that play important role in development.

###### a) Roles

Given below are the members of the project development lifecycle team and their essential skills [14]:

- **Stakeholder** represents interest groups whose needs must be satisfied by the project. It is a role that may be played by anyone who is (or potentially will be) materially affected by the outcome of the project.
- **Requirement Engineer:** He is responsible for eliciting the requirements from the users.
- **Analyst** represents customer and end-user concerns by gathering input from stakeholders to understand the problem to be solved and by capturing and setting priorities for requirements.
- **Architect** is responsible for designing the software architecture, which includes making the key technical

decisions that constrain the overall design and implementation of the project.

- **Developer** is responsible for developing a part of the system, including designing it to fit into the architecture, and then implementing, unit-testing, and integrating the components that are part of the solution.
- **Tester** is responsible for the core activities of the test effort, such as identifying, defining, implementing, and conducting the necessary tests, as well as logging the outcomes of the testing and analyzing the results.
- **Project Manager** leads the planning of the project in collaboration with stakeholders and team, coordinates interactions with the stakeholders, and keeps the project team focused on meeting the project objectives.
- Any **Role** represents anyone on the team that can perform general tasks.

##### b) Disciplines (Main Phases)

The BNF software development lifecycle process content comprises of the following disciplines which build a framework for the model:

- **Project Cost-Time Plan Estimation:** This phase is responsible for estimating the schedule and cost for developing the software product.
- **Requirement Generation:** The Requirement Generation phase is responsible for gathering a complete and executable set of requirements set for developing the product.
- **Design & Validation:** This phase uses the requirements elicited in the above phase to produce the preliminary and detailed design.
- **Code & Test:** In this phase the design is translated into a code and is tested for errors.

Every phase of BNF is accompanied by a testing phase to ensure quality assurance. Errors are detected and removed in the phase in which they get injected. The sooner they are detected the lower is the cost to remove them and the lesser is their affect on successive stages.

##### c) Tasks

A task is unit of work a role may be asked to perform [14]. In BNF the above mentioned disciplines are decomposed into 11 tasks or sub phases which are assigned to the above mentioned roles. These tasks include:

- **Requirement Elicitation:** Gathers requirements from the intended set of users.

- Requirement Analysis & Prioritization: Requirements elicited in the above activity are analyzed and prioritized here for implementation.
- Requirement Verification: The above list of prioritized requirements is read to the user for acceptance on the finalized list of requirements.
- Requirement Change Management: Changes in requirements that surface later while development are analyzed from system's perspective and are accommodated in this phase.
- Requirement Validation: Requirements are tested for execution using test cases.
- Conceptual Design: Conceptual or preliminary design of the software is built.
- Detailed Design: A detailed design of the software is built.
- Design Validation: Design is validated to detect design errors if any.
- Coding: Design built in the above phase is translated using a programming language.
- Unit Testing: Units of the software are tested for errors.
- Integration Testing: Modules are integrated and then tested after integration.

#### d) Artifacts

An artifact is something that is produced, modified, or used by a task. Roles are responsible for creating and updating artifacts. Artifacts produced by respective tasks include:

- Requirement Specification document
- Requirement verification checklists
- Traceability matrix
- State diagrams
- Sequence diagrams
- Fault tree analysis diagrams
- Code
- Unit test plans
- Integration test plans

These artifacts are subject to changes that might surface during later phases of software development generating different versions of these artifacts during actual development.

#### 2) BNF Process Content

Process content takes the method elements and relates them into an organized sequence of events as they occur.

Disciplines are made by organizing tasks from the method content) into activities which are applied in the respective discipline. Changes in gathered requirements are inherent with development of a product. To accommodate these changes, BNF encourages development to loop back to any previous phase. BNF process is iterative in nature that allows movement from one phase to any other required phase backward in the lifecycle to make changes and follow the sequential order if no other change surfaces. The BNF process has been illustrated graphically in Figure 1 below.

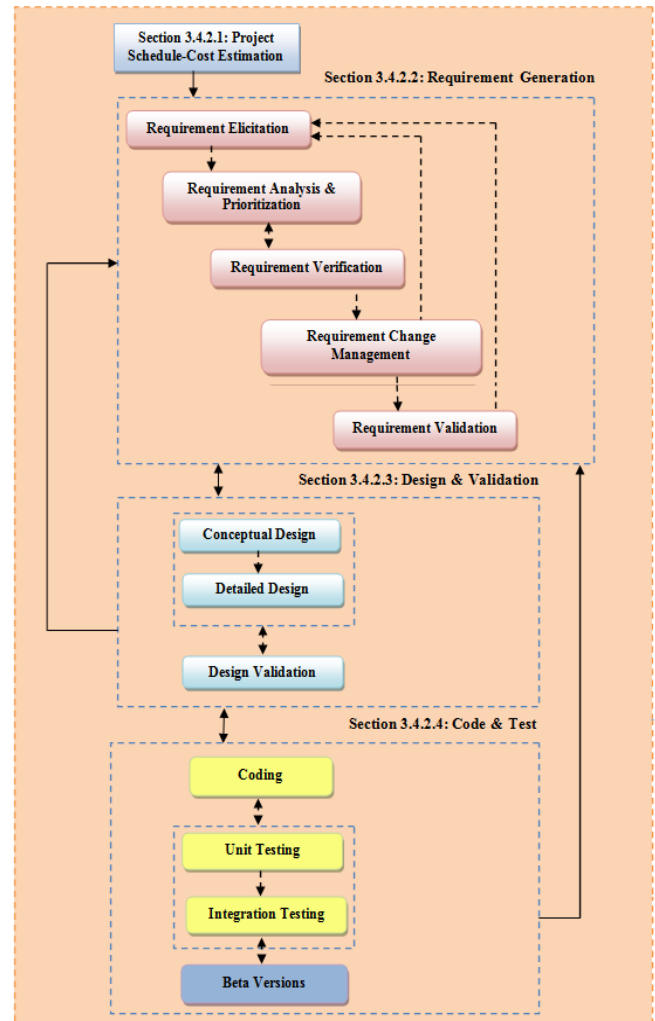


Figure 1. BNF Software Process Model

BNF uses a two dimensional approach to development methodology namely, the Method Content dimension and the Process Content Dimension. Both dimensions are inseparable and work hand in hand for software

development. These two dimensions and their relationship is illustrated in tables 1, 2 and 3 respectively.

Given below are the tables that map the relationship between the two dimensions.

**Table 1 Method-Process Content Table (Roles)**

	Method Content (Roles)						
	Stakeholder	Req. Engineer	System Analyst	Designer	Developer	Tester	Project Manager
Process Content	Req. Elicitation	✓	✓	✓			
	Req. Analysis & Prioritization			✓			
	Req. Verification		✓	✓			✓
	Req. Change Management	✓	✓				
	Req. Validation					✓	
	Architectural Design			✓			
	Conceptual Design			✓			✓
	Detailed Design			✓			
	Design Validation			✓			
	Coding				✓		
	Unit Testing					✓	
	Integration Testing					✓	
	Deployment				✓		✓

**Table 2 Method Process Content Table (Artifacts)**

	Method Content (Artifacts)								
	Req. Draft	Req. Doc	SRS	Trace Matrix	Test Cases	Design Doc	FTA	Code	Test Plan
Process Content	Req. Elicitation	✓							
	Req. Analysis & Prioritization	✓							
	Req. Verification		✓						
	Req. Change Management		✓	✓					
	Req. Validation				✓				
	Conceptual Design					✓			
	Detailed Design					✓			
	Design Validation					✓	✓		
	Coding							✓	
	Unit Testing								
	Integration Testing								✓
	Deployment								

**Table 3 Method Process Content Table (Disciplines)**

	Method Content (Disciplines)		
	Req. Generation	Design	Implementation
Process Content	Req. Elicitation	✓	
	Req. Analysis & Prioritization	✓	
	Req. Verification	✓	
	Req. Change Management	✓	
	Req. Validation	✓	
	Conceptual Design		✓
	Detailed Design		✓
	Design Validation		✓
	Coding		✓
	Unit Testing		✓
	Integration Testing		✓
	Deployment		✓

#### IV. CONCLUSION

The proposed BNF is a software development process model which can be manifested as a complete framework for the development of software. Unlike other software process models it not just explains the sequence in which the development activities are to proceed but also explains in detail the methodology that contains some specific disciplines, roles and artifacts necessary for successful development of the software product.

In practice, changes form an inherent part of a software development process and they emerge more with increase in system knowledge. Linear development of software development models like the Waterfall Model cannot be followed as it freezes in a process model. Incremental emphasizes on requirement prioritization of already acquired requirements while Spiral model emphasizes on risk management. Unlike other software process models BNF does not constrain all requirements to be complete during the requirement generation phase. Instead it elicits as many requirements as possible and lets development proceed sequentially as long as no new requirements are identified. As new requirements surface the development loops back to any other desired phase where changes are need to be accommodated. Hence BNF can be employed in software domains where requirements are subjective to many changes.

#### V. FUTURE WORK

Our future work involves validating our software process model on software development of an organization and comparing results of development with those of other software process models. One method for BNF model validation will be a case study results from actual implementation of BNF model. Second method includes a simulation that will compare the BNF model with other

models by simulating a software development work environment.

#### REFERENCES

- [1] Ian Sommerville, "Software Engineering", 8th Edition, 2006, pp. 89.
- [2] R.S. Pressman, "Software Engineering, A Practitioner's Approach", 5th ed. New York: McGraw-Hill, 2001, pp. 26.
- [3] B.W. Boehm, "Anchoring the Software Process", IEEE, IEEE Software, vol. 13, issue 4, July 1996, pp. 73-83.
- [4] E. Carmel, S. Becker, "A Process Model for Packaged Software Development", IEEE, IEEE Transaction on Engineering Management, vol. 42, Feb 1995, pp. 50-58.
- [5] W.W. Royce, "Managing the Development of Large Software Systems: Concepts and Techniques", IEEE, IEEE Computer Society, August 1970, pp. 1-9
- [6] B.W. Boehm, "A Spiral Model for Software Development and Enhancement", IEEE, IEEE Computer Society, vol. 21, issue 5, May 1988, pp. 61 – 72.
- [7] R.J. Madachy, "Software Process Dynamics", New Jersey: Willey Interscience, 2007, pp. 31.
- [8] R.S. Pressman, "Software Engineering, A Practitioner's Approach", 5th ed. New York: McGraw-Hill, 2001, pp. 32.
- [9] E.I. May, B. A. Zimmer, "The Evolutionary Development Model for Software", Hewlett-Packard Journal, Article 4, August 1996, pp. 1-8.
- [10] R.J. Madachy, "Software Process Dynamics", New Jersey: Willey Interscience, 2007, pp. 33.
- [11] R.S. Pressman, "Software Engineering, A Practitioner's Approach", 5th ed. New York: McGraw-Hill, 2001, pp. 34.
- [12] P. Kruchten, "Rational Unified Process Best Practices for Software Development Teams", Canada: rational Software, 2001.
- [13] Jeff Tian, Southern Methodist University, Dallas, "Software Quality Engineering", IEEE, Computer Society Publication, Willy InterScience, 2005.
- [14] Mahmood Khan Niazi. "Improving the Requirement Engineering Process through the Application of a key Process Areas Approach". AWRE'2002, pp.125-139.
- [15] Farooque Azam, Hina Gull, Saeeda Bibi and Sameera Amjad. "Bacck & Forth Software Process Model", Proceedings of 2<sup>nd</sup> International International Conference on Computer Engineering and Applications, pp.426-430, 9 March, 2009.
- [16] Article in a conference proceedings:
- [17] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.

# Performance Comparison of Block Truncation Coding based Image Retrieval Techniques using Assorted Color Spaces

Dr. H.B.Kekre  
Senior Professor  
Computer Engg. Dept.  
MPSTME, SVKM's NMIMS  
University,  
Mumbai, India  
hbkekke@yahoo.com

Sudeep D.Thepade  
Ph.D.Research Scholar &  
Associate Professor  
Computer Engg. Dept.  
MPSTME, SVKM's NMIMS  
University  
Mumbai, India  
sudeepthepade@gmail.com

Varun Lodha, Pooja  
Luthra, Ajoy Joseph,  
Chitrangada Nemani  
B.Tech (IT) Students  
MPSTME, SVKM's NMIMS  
University  
Mumbai, Ind

**Abstract**— *The paper presents exhaustive performance comparison of image retrieval techniques based on block truncation coding (BTC) using assorted color spaces. Including RGB total ten color spaces are considered for applying BTC to extract the feature vector in CBIR techniques. Further the image tiling is added to get three CBIR techniques per color space. In all performance comparison of thirty image retrieval techniques is done with the help of generic image database having 1000 images spread across 11 categories. For each proposed CBIR technique 55 queries (5 per category) are fired on the generic image database. To compare the performance of image retrieval techniques average precision and recall are computed of all queries. The results have shown the performance improvement (higher precision and recall values) with proposed color-BTC methods compared to gray-BTC in all color spaces except 'rgb' color space. Image tiling does not help to improve the performance in the chromaticity-luminance based color spaces (Kekre's LUV, YCbCr, YUV, YIQ, Kekre's YCgCb), while it helps in non-luminance color spaces (RGB, HSV, XYZ, HIS). Overall Kekre's LUV color space based BTC gives best performance in image retrieval.*

**Keywords** — *CBIR, BTC, Color Space, Image Tiling, VQ, RGB, HSV, XYZ, HIS, rgb, Kekre's LUV, YCbCr, YUV, YIQ, Kekre's YCgCb*

## I. INTRODUCTION

The large numbers of images are being generated from a variety of sources (digital camera, digital video, scanner, the internet etc.) which have posed technical challenges to computer systems to store/transmit and index/manage image data effectively to make such collections easily accessible. Image compression deals with the challenge of storage and transmission, where significant advancements have been made [1,4,5]. The challenge to image indexing is studied in the context of image database [2,6,7,10,11], which has become one of the promising and important research area for researchers from a wide range of disciplines like computer vision, image processing and database areas. The thirst of

better and faster image retrieval techniques is increasing day by day. Some of important applications for CBIR technology could be identified as art galleries [12,14], museums, archaeology [3], architecture design [8,13], geographic information systems [5], weather forecast [5,22], medical imaging [5,18], trademark databases [21,23], criminal investigations [24,25], image search on the Internet [9,19,20].

## A. Content Based Image Retrieval

In literature the term content based image retrieval (CBIR) has been used for the first time by Kato et.al. [4], to describe his experiments into automatic retrieval of images from a database by colour and shape feature. The typical CBIR system performs two major tasks [16,17]. The first one is feature extraction (FE), where a set of features, called feature vector, is generated to accurately represent the content of each image in the database. The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their feature vectors is used to retrieve the top "closest" images [16,17,26]. For feature extraction in CBIR there are mainly two approaches [5] feature extraction in spatial domain and feature extraction in transform domain. The feature extraction in spatial domain includes the CBIR techniques based on histograms [5], BTC [1,2,16], VQ [21,25,26]. The transform domain methods are widely used in image compression, as they give high energy compaction in transformed image [17,24]. So it is obvious to use images in transformed domain for feature extraction in CBIR [23]. But taking transform of image is time consuming and also needs all images of database to be of same size to get similar feature vectors. This limitation is overcome in CBIR using Block Truncation Coding (BTC).

## B. Block Truncation Coding

Block truncation coding (BTC) is a relatively simple image coding technique developed in the early years of digital imaging more than 29 years ago. Although it is a simple technique, BTC has played an important role in the history of



digital image coding in the sense that many advanced coding techniques have been developed based on BTC or inspired by the success of BTC [1,2]. Block Truncation Coding (BTC) was first developed in 1979 for grayscale image coding [2]. This method first divides the image to be coded into small non-overlapping image blocks (typically of size  $4 \times 4$  pixels to achieve reasonable quality). The small blocks are coded one at a time. For each block, the original pixels within the block are coded using a binary bit-map the same Upper Mean Color size as the original blocks and two mean pixel values. In the original implementation the block mean and the variance of the pixels are used to preserve the first and second moment of the blocks. The descriptors here follow a later version of BTC, which was shown to give better performance [1,16]. The method first computes the mean pixel value of the whole block and then each pixel in that block is compared to the block mean. If a pixel is greater than or equal to the block mean, the corresponding pixel position of the bitmap will have a value of 1 otherwise it will have a value of 0. Two mean pixel values one for the pixels greater than or equal to the block mean and the other for the pixels smaller than the block mean are also calculated. At decoding stage, the small blocks are decoded one at a time. For each block, the pixel positions where the corresponding bitmap has a value of 1 is replaced by one mean pixel value and those pixel positions where the corresponding bitmap has a value of 0 is replaced by another mean pixel value. It was quite natural to extend BTC to multi spectrum images such as color images. Most color images are recorded in RGB space, which is perhaps the most well-known color space. As described previously, BTC divides the image to be coded into small blocks and code them one at a time. For single bitmap BTC of color image, a single binary bitmap the same size as the block is created and two colors are computed to approximate the pixels within the block. To create a binary bitmap in the RGB space, an inter band average image (IBAI) is first created and a single scalar value is found as the threshold value. The bitmap is then created by comparing the pixels in the IBAI with the threshold value [16,20].

## II. CONSIDERED COLOR SPACES

The considered ten color spaces and their inter-conversion formulae are given below.

### A. HSV Color Space

The HSV [27] stands for the Hue, Saturation, and Value based on the artists (Tint, Shade, and Tone). The Value represents intensity of a color, which is decoupled from the color information in the represented image. The Hue and Saturation components are intimately related to the way human eye perceives color resulting in image processing algorithms with physiological basis.

### B. XYZ Color Space

The Y primary is intentionally defined to match closely to luminance, while X and Z primaries give color information. The main advantage of the CIE XYZ space [27,28] is that this space is completely device-independent.

### C. rgb Color Space

In order to eliminate the influence of illumination intensity, color information (R,G and B) can be normalised to get rgb color space[29,30]. The transformation from RGB space to rgb space a very simple color normalization method, whose advantage is the quick computation.

### D. HSI Color Space

To convert RGB to HSI, first RGB need to be converted into normalized one as rgb [29,30] and then rgb is converted to HSI.

### E. Kekre's LUV Color Space

Kekre's LUV color Space [1,12,13,14], which is special case of Kekre Transform. Here L gives luminance and U and V gives chromaticity values of color image. Positive value of U indicates prominence of red component in color image and negative value of V indicates prominence of green component. This needs the conversion of RGB to LUV components.

### F. YCbCr Color Space

In YCbCr color Space [12, 20], Y gives luminance and Cb and Cr give chromaticity values of color image. To get YCbCr components we need the conversion of RGB to YCbCr components.

### G. YUV Color Space

In YUV color Space [16], the YUV model defines a color space in terms of one luminance (brightness) and two chrominance (color) components. The YUV color model is used in the PAL, NTSC, and SECAM composite color video standards. Previous black-and-white systems used only luminance (Y) information and color information (U and V) was added so that a black-and-white receiver would still be able to display a color picture as a normal black and white pictures. YUV models human perception of color in a different way than the standard RGB model used in computer graphics hardware. The human eye has fairly little color sensitivity: the accuracy of the brightness information of the luminance channel has far more impact on the image discerned than that of the other two.

### H. YIQ Color Space

The YIQ color space [28, 29] is derived from YUV color space and is optionally used by the NTSC composite color video standard. The "I" stands for in phase and "Q" for quadrature, which is the modulation method used to transmit the color information.

### I. Kekre's YCgCb Color Space

The novel color space Kekre's YCgCb is introduced in [27,28], where Y is luminance and Cg and Cb are chromaticity values.

## III. IMAGE TILING

Image Tiling [18,24,31] means dividing any given image into non-overlapping cells or fragments. The size of each tile is

such that it divides the image into N equal parts and also keeping the size of each tile the same.

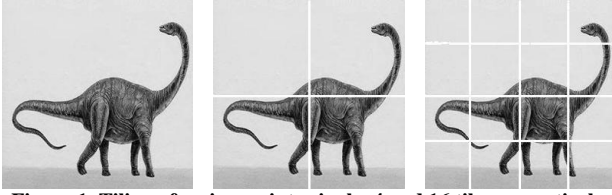


Figure 1. Tiling of an image into single, 4 and 16 tiles respectively

#### IV. CBIR USING COLOR BLOCK TRUNCATION CODING

In original BTC [1,2] we divide the image into R, B, and G components and compute the inter band average image (IBAI) which is the average of all the components(R, G, and B) and mean of interband average image is taken as threshold. But the disadvantage of this method is that if one of the component is prominent than the other component then that component dominates the threshold value, reducing the effect of other component. A more general approach is by using three independent R, G and B components of color images to calculate three different thresholds and then apply BTC to each individual R, G and B planes. Let the thresholds be TR, TG and TB, which could be computed as per the equations given below

$$TR = \frac{1}{m*n} \sum_{i=1}^m \sum_{j=1}^n R(i,j) \quad (1)$$

$$TG = \frac{1}{m*n} \sum_{i=1}^m \sum_{j=1}^n G(i,j) \quad (2)$$

$$TB = \frac{1}{m*n} \sum_{i=1}^m \sum_{j=1}^n B(i,j) \quad (3)$$

Here three binary bitmaps will be computed as BMr, BMg and Bmb. If a pixel in each component (R, G, and B) is greater than or equal to the respective threshold, the corresponding pixel position of the bitmap will have a value of 1 otherwise it will have a value of 0.

$$BMr(i,j) = \begin{cases} 1, \text{if } \dots R(i,j) \geq TR \\ 0, \dots \text{if } \dots R(i,j) < TR \end{cases} \quad (4)$$

$$BMg(i,j) = \begin{cases} 1, \text{if } \dots G(i,j) \geq TG \\ 0, \dots \text{if } \dots G(i,j) < TG \end{cases} \quad (5)$$

$$Bmb(i,j) = \begin{cases} 1, \text{if } \dots B(i,j) \geq TB \\ 0, \dots \text{if } \dots B(i,j) < TB \end{cases} \quad (6)$$

Two mean colors one for the pixels greater than or equal to the threshold and other for the pixels smaller than the threshold are also calculated [16]. The upper mean color UM (Rm1, Gm1, Bm1) is given as follows.

$$Rm1 = \frac{1}{\sum_{i=1}^m \sum_{j=1}^n BMr(i,j)} * \sum_{i=1}^m \sum_{j=1}^n BMr(i,j) * R(i,j) \quad (7)$$

$$Gm1 = \frac{1}{\sum_{i=1}^m \sum_{j=1}^n BMg(i,j)} * \sum_{i=1}^m \sum_{j=1}^n BMg(i,j) * G(i,j) \quad (8)$$

$$Bm1 = \frac{1}{\sum_{i=1}^m \sum_{j=1}^n Bmb(i,j)} * \sum_{i=1}^m \sum_{j=1}^n Bmb(i,j) * B(i,j) \quad (9)$$

And the Lower Mean LM= (Rm2, Gm2, Bm2) is computed as following equations

$$Rm2 = \frac{1}{m*n - \sum_{i=1}^m \sum_{j=1}^n BMr(i,j)} * \sum_{i=1}^m \sum_{j=1}^n \{1 - BMr(i,j)\} * R(i,j) \quad (10)$$

$$Gm2 = \frac{1}{m*n - \sum_{i=1}^m \sum_{j=1}^n BMg(i,j)} * \sum_{i=1}^m \sum_{j=1}^n \{1 - BMg(i,j)\} * G(i,j) \quad (11)$$

$$Bm2 = \frac{1}{m*n - \sum_{i=1}^m \sum_{j=1}^n Bmb(i,j)} * \sum_{i=1}^m \sum_{j=1}^n \{1 - Bmb(i,j)\} * B(i,j) \quad (12)$$

These Upper Mean and Lower Mean together will form a feature vector or signature of the image. For every image stored in the database these feature vectors are computed and stored in feature vector table. In case of other color spaces firstly the conversion of RGB to Color-components of considered color space is done and then the same color-BTC technique is applied to color-components for feature extraction from images. Thus the feature vectors are formed using color-components of respective color spaces. For image tiling based color-BTC, the process of BTC is applied to all image tiles separately to obtain the feature vectors of image. Thus per tile six BTC color descriptors can be obtained and put into feature vector.

#### V. IMPLEMENTATION

The implementation of the three CBIR techniques is done in MATLAB 7.0 using a computer with Intel Core 2 Duo Processor T8100 (2.1GHz) and 2 GB RAM. The CBIR techniques are tested on the image database [15] of 1000 variable size images spread across 11 categories of human being, animals, natural scenery and manmade things. The categories and distribution of the images are Tribes (85), Buses(99), Beaches(99), Dinosaurs(99), Elephants(99), Roses(99), Horses(99), Mountains(61), Airplanes(100), Monuments(99) and Sunrise (61). Figure 2 gives the sample images from generic image database.



Figure 2. Sample Images from Generic Image Database[32]  
[Image database contains total 1000 images with 11 categories]

To assess the retrieval effectiveness, we have used the precision and recall as statistical comparison parameters [1,2] for the proposed CBIR techniques. The standard definitions for these two measures are given by following equations.

$$\text{Precision} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_images\_retrieved}} \quad (13)$$

$$\text{Recall} = \frac{\text{Number\_of\_relevant\_images\_retrieved}}{\text{Total\_number\_of\_relevant\_images\_in\_database}} \quad (14)$$

## VI. RESULTS AND DISCUSSION

For testing the performance of each proposed CBIR technique first five images from each category are fired on the database as queries. The average precision and average recall are computed by grouping the number of retrieved images.

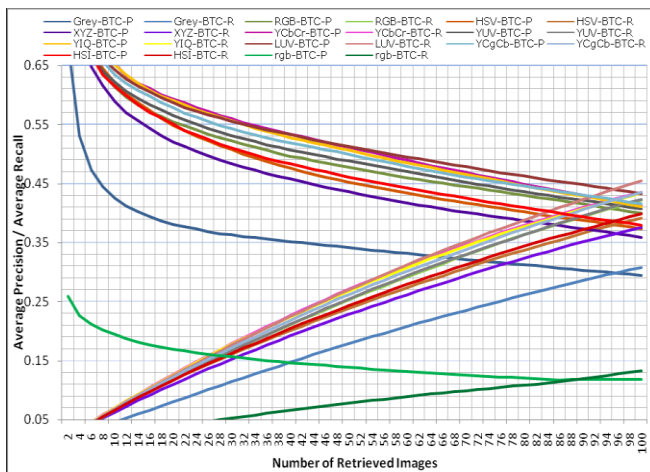


Figure 3. Average Precision/Recall plotted against Number of retrieved images for various color spaces using complete image (1-Tile)

Figure 3 gives the average precision and average recall values plotted against number of retrieved images for the BTC-CBIR techniques applied on complete image (1-Tile). From the figure it can be noted that by means of higher precision and recall values, all color-BTC techniques are giving better image retrieval than gray-BTC except „rgb“ color space. Kekre's LUV color space is giving best performance while YCbCr and Kekre's YCgCb is second best in performance.

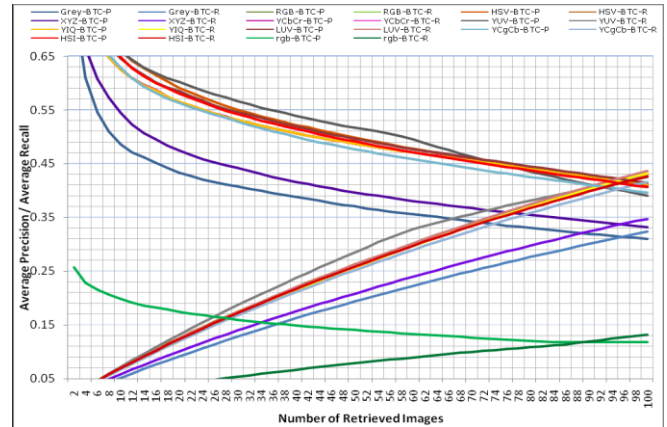


Figure 4. Average Precision/Recall plotted against Number of retrieved images for various color spaces using 4-Tiles of image

Average precision and average recall values of applying BTC-CBIR techniques with 4-Tiles of image considered are given in figure 4. Here also Kekre's LUV based BTC-CBIR is outperforming other color spaces, while second best performance is shown by YIQ color space. Even in 4 tiling based BTC-rgb color space continues to give worst performance. The trend that color-BTC outperforms gray-BTC is continued even in 4 tiling.

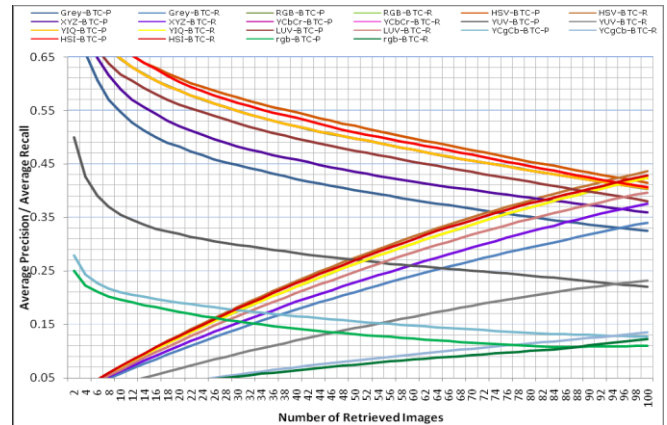


Figure 5 gives the average precision and average recall values plotted against number of retrieved images for the BTC-CBIR techniques applied on 16 tiles of image. Here HSV color space gives best performance followed by performance by HSI and YIQ color spaces. Kekre's LUV color space which gave best image retrieval in 1tile and 4 tile based BTC stands fourth here in case of performance. The rgb color space continues to give worst performance. The crossover point values of average precision and average recall curves can be used as the performance indicator for image retrieval techniques [23]. Higher the crossover point value better the performance of image retrieval technique is. The comparison of performance of discussed BTC-CBIR techniques is shown in figure 6 and figure 7.

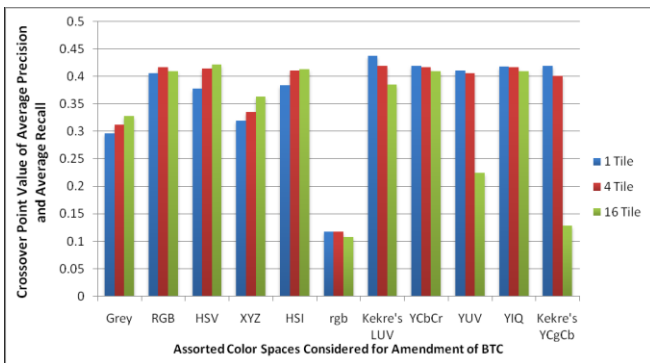


Figure 6. Comparison of performance of BTC-CBIR using image tiling techniques for respective color spaces

Figure 6 gives performance comparison of BTC-CBIR techniques based on image tiling for respective color spaces. Along with in gray-BTC, the performance of image retrieval improves with image tiling even in non-luminance color spaces like RGB, HSV, XYZ, HSI. Image tiling does not help in improving the performance of BTC based image retrieval in case of luminance-chromaticity based color spaces like Kekre's LUV, YCbCr, YUV, YIQ and Kekre's YCgCb.

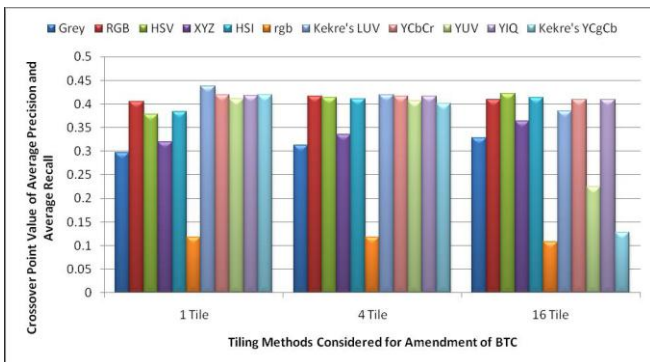


Figure 7. Comparison of performance of BTC-CBIR using assorted color spaces for respective image tiling techniques

The performance comparison based on image tiling methods for BTC-CBIR techniques using assorted color spaces is shown in figure 7. Here in 1 tile based BTC-CBIR, it can be observed that the performance of luminance-chromaticity based color spaces (Kekre's LUV, YCbCr, YIQ, YUV, Kekre's YCgCb) is higher than other color spaces. In 4-tiles the performance of HSV and HSI becomes equal to that of luminance-chromaticity based color spaces. While as the image tiling is increased to 16 tiles, the performance of CBIR techniques based on non-luminance color spaces shows improvement. Overall rgb gives worst performance and Kekre's LUV gives the best.

## VII. CONCLUSIONS

So far many CBIR techniques have been proposed, but till the researchers are craving for better and faster image retrieval solutions. The paper presented the exhaustive comparison of image retrieval techniques based on color block truncation coding (color-BTC) using 10 assorted color spaces. The

techniques are tested on image database of 1000 generic images spread across 11 different categories. It is observed that all color spaces except „rgb“ give performance improvement as color-BTC compared to gray-BTC for image retrieval. Image tiling helps to improve the performance further only in case of non-luminance color spaces (RGB, HSV, HSI and XYZ) but not in luminance-chromaticity based color spaces. Overall Kekre's LUV color space proves to be the best with no image tiling for BTC-CBIR.

## IX. REFERENCES

- [1] H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, Number 3, pp. 172-180, Summer 2008. Available online at <http://www.waset.org/ijecse/v2/v2-3-23.pdf>
- [2] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigues College of Engg., Mumbai. Is uploaded on online ACM portal.
- [3] H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html), Volume 1, No. A08, pp. 31-46, Autumn 2008.
- [4] Hirata K. and Kato T. "Query by visual example – content-based image retrieval", In Proc. of Third International Conference on Extending Database Technology, EDBT'92, 1992, pp 56-71
- [5] H.B.Kekre, Sudeep D. Thepade, "Rendering Futuristic Image Retrieval System", National Conference on Enhancements in Computer, Communication and Information Technology, EC2IT-2009, 20-21 Mar 2009, K.J.Somaiya College of Engineering, Vidyavihar, Mumbai-77.
- [6] Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Transactions On Image Processing, Volume 11, Number 2, pp.146-158, February 2002.
- [7] B.G.Prasad, K.K. Biswas, and S. K. Gupta, "Region –based image retrieval using integrated color, shape, and location index", International Journal on Computer Vision and Image Understanding Special Issue: Colour for Image Indexing and Retrieval, Volume 94, Issues 1-3, April-June 2004, pp.193-233.
- [8] H.B.Kekre, Sudeep D. Thepade, "Creating the Color Panoramic View using Medley of Grayscale and Color Partial Images ", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No. 3, Summer 2008. Available online at [www.waset.org/ijecse/v2/v2-3-26.pdf](http://www.waset.org/ijecse/v2/v2-3-26.pdf).
- [9] Stian Edvardsen, "Classification of Images using color, CBIR Distance Measures and Genetic Programming", Ph.D. Thesis, Master of science in Informatics, Norwegian university of science and Technology, Department of computer and Information science, June 2006.
- [10] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "DCT Applied to Row Mean and Column Vectors in Fingerprint Identification", In Proceedings of International Conference on Computer Networks and Security (ICCNS), 27-28 Sept. 2008, VIT, Pune.
- [11] Zhibin Pan, Kotani K., Ohmi T., "Enhanced fast encoding method for vector quantization by finding an optimally-ordered Walsh transform kernel", ICIP 2005, IEEE International Conference, Volume 1, pp 1 - 573-6, Sept. 2005.
- [12] H.B.kekre, Sudeep D. Thepade, "Improving „Color to Gray and Back“ using Kekre's LUV Color Space", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009. Is uploaded and available online at IEEE Xplore.
- [13] H.B.Kekre, Sudeep D. Thepade, "Image Blending in Vista Creation using Kekre's LUV Color Space", SPIT-IEEE Colloquium and International Conference, Sardar Patel Institute of Technology, Andheri, Mumbai, 04-05 Feb 2008.
- [14] H.B.Kekre, Sudeep D. Thepade, "Color Traits Transfer to Grayscale Images", In Proc.of IEEE First International Conference on Emerging



- Trends in Engg. & Technology, (ICETET-08), G.H.Raisoni COE, Nagpur, INDIA. Uploaded on online IEEE Xplore.
- [15] <http://wang.ist.psu.edu/docs/related/Image.org> (Last referred on 23 Sept 2008)
- [16] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.
- [17] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS), Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at [www.IJCSNS.org](http://www.IJCSNS.org).
- [18] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Walsh Transform over Row Mean and Column Mean using Image Fragmentation and Energy Compaction for Image Retrieval", International Journal on Computer Science and Engineering (IJCE), Volume 2S, Issue1, January 2010, (ISSN: 0975-3397). Available online at [www.enggjournals.com/ijcse](http://www.enggjournals.com/ijcse).
- [19] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue I, Feb.2010, pp.9-18, Available online [www.icgst.com/gvip/Volume10/Issue1/P1150938876.html](http://www.icgst.com/gvip/Volume10/Issue1/P1150938876.html)
- [20] H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 2-14. Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html) (ISSN: 0974-0627).
- [21] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 55-65. Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html) (ISSN: 0974-0627).
- [22] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Performance Comparison for Face Recognition using PCA, DCT & Walsh Transform of Row Mean and Column Mean", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue II, Jun.2010, pp.9-18, Available online <http://209.61.248.177/gvip/Volume10/Issue2/P1181012028.pdf>.
- [23] H.B.Kekre, Sudeep D. Thepade, "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image", International Journal of Information Retrieval, Serials Publications, Volume 2, Issue 1, 2009, pp. 72-79 (ISSN: 0974-6285)
- [24] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.
- [25] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi, "Improved Texture Feature Based Image Retrieval using Kekre's Fast Codebook Generation Algorithm", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.
- [26] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval by Kekre's Transform Applied on Each Row of Walsh Transformed VQ Codebook", (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is invited at ICWET 2010. Also will be uploaded on online ACM Portal.
- [27] H.B.Kekre, Sudeep D. Thepade, Adib Parkar, "A Comparison of Kekre's Fast Search and Exhaustive Search for various Grid Sizes used for Colouring a Greyscale Image", 2nd International Conference on Signal Acquisition and Processing (ICSAP 2010), IACSIT, Bangalore, pp. 53-57, 9-10 Feb 2010, The paper is uploaded on online IEEE Xplore
- [28] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Adib Parkar, "Using Assorted Color Spaces and Pixel Window Sizes for Colorization of Grayscale Images", ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is uploaded on online ACM Portal.
- [29] Mark D. Fairchild, "Color Appearance Models", 2nd Edition, Wiley-IS&T, Chichester, UK, 2005. ISBN 0-470-01216-1
- [30] Rafael C. Gonzalez and Richard Eugene Woods "Digital Image Processing", 3rd edition, Prentice Hall, Upper Saddle River, NJ, 2008. ISBN 0-13-168728-X. pp. 407-413.S
- [31] H.Lu, B.Ooi and K.Tan. Efficient image retrieval by color contents. In Intl. Conf. on Applications of Databases, volume 819, pages 95-108, June 21-23 1994

#### X. AUTHOR BIOGRAPHIES



**Dr. H. B. Kekre** Kekre has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./B.Tech projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 300 papers in National / International Conferences and Journals to his credit. He was Senior Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE Recently ten students working under his guidance have received best paper awards. Currently 10 research scholars are pursuing Ph.D. program under his guidance.



**Sudeep D. Thepade** has Received B.E.(Computer) degree from North Maharashtra University with Distinction in 2003. M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, currently pursuing Ph.D. from SVKM's NMIMS, Mumbai. He has about 08 years of experience in teaching and industry. He was Lecturer in Dept. of Information Technology at Thadomal Shahani Engineering College, Bandra(w), Mumbai for nearly 04 years. After working as Assistant Professor now he is Associate Professor in Computer Engineering at Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. His areas of interest are Image Processing and Computer Networks. He has about 100 papers in National/International Conferences/Journals to his credit with a Best Paper Award at International Conference SSPCCIN-2008, Second Best Paper Award at ThinkQuest-2009 National Level paper presentation competition for faculty, second best research project award at MANSHODHAN-2010 and Best Paper Award at Springer International Conference ICCCT-2010.

**Chitrangada Nemani, Ajoy Joseph, Pooja Luthra, Varun Lodha** are currently pursuing B.Tech. (IT) from MPSTME, SVKM's NMIMS University, Mumbai. There areas of interest is Content Based Image Retrieval in Image Processing

# Development of a Project-Based Learning Approach in Requirement Engineering

Nor Azliana Akmal Jamaludin  
Fakulti Industri Teknologi Maklumat,  
Universiti Industri Selangor (UNISEL)  
45600 Bestari Jaya, Selangor, Malaysia  
norazliana@unisel.edu.my

Shamsul Sahibuddin  
Advanced Informatics School, UTM International  
Campus, Universiti Teknologi Malaysia  
Jalan Semarak, 54100 Kuala Lumpur, Malaysia  
shamsul@utm.my

Kamaruzaman Jusoff  
Faculty of Forestry,  
Universiti Putra Malaysia (UPM )  
43400 UPM Serdang Selangor, Malaysia  
kamaruz@putra.upm.edu.my

Nur Hafizah Hidayat  
Fakulti Industri Teknologi Maklumat,  
Universiti Industri Selangor (UNISEL)  
45600 Bestari Jaya, Selangor, Malaysia  
nur\_hafizah @unisel.edu.my

**Abstract**—Project failure is due to the lack of Requirement Engineering (RE) practice. The Industry needs to allocate another cost to send their employee for additional training before the employee can contribute to the job specification. It indicates that current practices of delivery mechanisms at the university fail to deliver graduates with employability skills. The objective of this research is to identify weaknesses in current practice of teaching Software Engineering undergraduate in Requirement Engineering. Additionally, this paper emphasized that Project-Based Learning (PjBL) is a right method for delivery mechanisms to enhance Software Engineering undergraduate skills particularly in RE. The PjBL is a superset to Problem-Based Learning, Individual-Collaborative Learning and Product-Based Learning. The intersection can strongly assist in the learning environment. Future work should be carried out to design the framework of PjBL, measuring the effectiveness of PjBL and the electronic Learning eNvironment (eLIN) system as a supportive tools to make PjBL successful.

**Keywords**—Software Engineering education; Project-Based Learning (PjBL); Requirement Engineering; Problem-Based Learning; Individual & Collaborative Problem Solving and Product-Based Learning.)

## I. INTRODUCTION

The number, size, and application domains of computer programs have grown dramatically. As a result, hundreds of billions are being spent on software development, and the livelihood and lives of most people depend on the effectiveness of this development. Since the dawn of computing in the 1940s, the use of computers and its applications have grown at a staggering rate. Software plays a central role in almost all aspects of our daily lives: in government, banking and finance, education, transportation, entertainment, medicine, agriculture, and law [37]. The current trend shows that students'

Cumulative Grade Point Average (CGPA) does not represent their skill [35]. In addition, academic supply does not meet the industry demands [17]. Recent trends with the effects of the recession, the unemployment rate is increasing. Because of this economic situation, something needs to be done for the students to prepare themselves before entering into the work force and these new graduates must understand what skills the industry is seeking [50]. Based on that reason, this research carried out the first objective that is to identify weaknesses in current practice of teaching Software Engineering undergraduate in Requirement Engineering.

## II. INDUSTRY AND SOFTWARE ENGINEERING EDUCATION PERSPECTIVE

### A. Industry Collaboration

The report shown in Table 1 marked a decrease in project success rates, with 32% of all projects surveyed were successful defined as projects delivered on time and within budget, with required features and functions. Another 44% of all projects surveyed were challenged defined as late, over budget, and/or with less than the required features and functions. The remaining 24% of all projects surveyed failed defined as cancelled prior to completion or delivery and never used [14]. In software development industry, companies will usually tackle three main problems: cost minimization, tight deadlines and quality product. An engineering approach in software development will help them to overcome these problems.



TABLE 1. THE TREND OVER THE PAST 15 YEARS BY STANDISH GROUP'S CHAOS (2009) REPORT.

Year	'09 (%)	'06 (%)	'04 (%)	'02 (%)	'00 (%)	'98 (%)	'96 (%)	'94 (%)
Successful Project	32	35	29	34	28	26	27	16
Challenged Projects	44	19	53	15	23	28	40	31
Failed Projects	24	46	18	51	49	46	33	53

Industry needs employees who are skilled in developing new application system, but they do not have the time or trainers to complete the training. It is supported by Bernhart et al (2006). He states that if students fail to deliver a minimum quality at the beginning, succeeding phases may receive a reduced quality as a work basis. This may reduce the learning experience in later phases [6]. Some researchers have concluded that businesses and universities share some similar challenges, and increased cooperation between the two entities will assist in shared solutions for both [15]. Therefore, student should be taught seriously in software development process as a field of Software Engineering [5, 25].

#### B. Future Challenges in Software Engineering Education

Boehm (2006) identifies future challenges in Software Engineering education with his simple hypothesis: "software people don't like to see Software Engineering done unsuccessfully, and try to make things better". Mead (2009) was look into the challenges from global reach of education, new creative evolution of delivery mechanism, new professional efforts and the need of engage in leadership in Software Engineering education [10, 13, 27, 39]. The study looked into the curriculum of undergraduate particularly in developing mission-critical system and non-critical system.

Furthermore, the continuous research is still looking for methods on how to maintain vitality and how to develop a new generation of Software Engineering educators through conference, working groups and committees electronic publishing [10, 39]. They also encourage others to join Software Engineering profession. Licensing or certification will become a major trend in the future and will create new specialization. However, looking at the fundamental concepts have not changed significantly since progress took shape some thirty years ago. As in hardware design, the technology evolves, but the concepts remain [2]. Students usually learn best by physically doing something. Students of course will be more interested to come to school if they think that education is fun and beneficial in the future and not as a burden [3].

#### C. Improvement should be included into Software Engineering Education

Many of the skills that students are expected to have can only be learned by doing hands-on practices. Furthermore, Jackson (2007) discussed on interacting with real customers on tight deadlines and budgets with high expectations, and being able to work effectively in an almost exclusively team

oriented environment with increasingly complex team structures and compositions. Software Engineering is an appropriate course to meld all of these ideas together to produce better simulates a real world environment [40]. It is because software products have helped people to be more efficient and productive such as Collaborative Application Lifecycle Management Solutions from IBM Rational (2010). It make more effective problem solvers, and it provide with an environment for work and play which is often safer, more flexible, and less confining. As a result, Software Engineering curriculum should involve five complementary elements [43] including principles practices, applications, tools and mathematics.

The researcher outlines the similarity of the technological characteristics from Engineering Education that needs to be adapted into Software Engineering Education to increase employability. The profiles may be conveniently sketched in terms of three components [42] that include knowledge, skills and attitudes. However, sustainable world-class performance will not occur if there is a misalignment between a university programmes objectives and actual market requirements. In addition, effective faculty wide coordination in relation to market driven initiatives is essential for ensuring the effective use of company resources [18]. Employers now focus on adaptation, cost reduction, increased productivity, and new markets, products and services.

### III. REQUIREMENT ENGINEERING

#### A. Requirement Engineering Challenges

Requirement Engineering (RE) practice appears to have improved collaboration ties and may have led to cultural changes that values cooperation, quality, and customer satisfaction [12]. Currently the success of a software system depends on how well it fits the needs of its users and its environment. Requirements encompass more than desired functionality – users increasingly demand systems that are usable, reliable, secure and economical, while product developers want to be able to adapt and enhance products rapidly, in response to changing both user's needs and environmental conditions [9]. Education become a kick-start in developing skills amongst the students (future employee) to cater the problem part (RE) before continues to solution part (other software artifacts such as design, development, testing, quality and maintenance) to improved productivity, quality and risk management. This is because customer satisfaction is the most pertinent priority in producing the successful product. Zave (1997) provides one of the clearest definitions of RE: "Requirements Engineering is the branch of Software Engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behavior, and to their evolution over time and across software families".

#### IV. PROJECT-BASED LEARNING

##### A. *The Comparative Study of Universities Experience in using PjBL*

Hence, a few experiences from universities with a variety of domain that implemented project-based works in their delivery mechanism are discussed. Firstly, Technological Education Institute of Larissa, Greece, use PjBL to allow students to improve their cognitive, collaborative, methodological and technical skills on Enterprise Resource Planning (ERP) systems through active participation in challenging and interdisciplinary ERP projects [21].

Secondly, Oregon State University in general, analyse and synthesize the collaborative and Project-Based Learning features resulted in 32 design features that were placed in the following six categories: (i) learning group size, (ii) functional spaces for learning activities, (iii) adjacencies, (iv) furnishings, (v) psychological and physiological support of the learners, and (vi) structural aspects. It can be used for architect, educator, planner and learner [51].

Thirdly, Department of Computer Engineering, Boğaziçi University in subject Cmpe450 previously used lecture oriented course, small group projects were assigned to students, in expectation that they should apply the theory to practice and they were expected to submit well-documented findings. Then, they changed to emphasize the project in terms of its size, scope, complexity and grading. The project topics included an e-learning system, various versions of department automation portal and a purchase automation system for the whole university. The first demonstrations were total disasters. It is after the first demonstration that they realize the severity of the expectations and what failure means. Therefore the documentation should be very precise to enable hiring of an administrator for the final product. It is too hard and often not applicable to map the theory and the practice part of the course [49].

Fourthly, Rowan University in Glassboro, NJ and Fairfield University in Fairfield, CT is introducing PjBL in their theoretical Software Engineering course. Local-Remote Team Organization and Communication Techniques (represent PjBL tool) was used in the theoretical Software Engineering course. Three of the projects were to extend existing products and one project was completely new. This research argues that students can be prepared to effectively join industry and keep the US technology workforce competitive through a curriculum that includes a theoretical Software Engineering course with real-world projects and the collaboration of paired teams across two or more universities. It was also appealing that all team members, local and remote, could have instant access to their materials from anywhere [44].

Fifthly, College of Arts and Sciences (Applied Science) Universiti Utara Malaysia called the method as Blended Project Based Learning. In support of the use of PjBL, a prototype known as md-Matrix has been used to assist students in choosing an appropriate development methodology based on their project requirements. The proposed decision tool can be considered as necessary to assist the students (novice developers) in choosing an appropriate development

methodology for their project. The implementation of the tool in blended PjBL will not only make the decision making process more effective but also help them with better learning [45].

Sixthly, Department of Mechanical Engineering, Politeknik Kota Bharu, tested Project-Based Learning module based on socio-constructivist approach in Diploma in Automotive. The project development ran a pilot test within six month to explore the influence of the utilization of the PjBL module on students' meta-cognition, motivation and self-regulation. In addition, it transforms teaching from 'lecturers/supervisors telling' to 'engineering students doing'. The result shows that the overall engineering students reported higher level of motivation, self-confidence and self-regulation throughout the process. The PjBL also enhances the cognitive and critical thinking in engineering problem solving among students. The use of PjBL module enables students to plan their project easily, work collaboratively with peers with minimal supervision from supervisors or lecturers and successfully complete their project on time [23].

Seventhly, Civil engineering degree at University of Malaya, Engineering Surveying Camp (Year-I of a four-year) test Outcome Based Education Using Project Based Learning in two-week field course. The aims are to evaluate the effectiveness, and to identify potential improvement. The need to change the paradigm of the majority of lecturers that are used to the conventional teaching methods, which worked on were real project to be completed within a short duration of the course. The students were primarily focused on the achievement of course outcomes. The academics, though, had another goal of ensuring that the project was completed with quality outputs to ensure the other aspects including the day to day running of the camp, transportation, logistics, health and safety on top of the course outcomes. The technical aspects were achievable though with lesser degree for the engineering design. On soft skills, students demonstrated an overall improvement of competency but it was difficult to ascertain the levels for the average students while the best and poor performers were easily observed [24].

##### B. *Project-Based Learning Method in Requirement Engineering Education*

Experiential method refers to the terms of evaluation or what has changed or improved as a result. Because of that, Project-Based Learning (PjBL) is a model that organizes learning around projects. According to the definitions found in PjBL handbooks for teachers, projects are complex tasks, based on challenging questions or problems, that involves students in designing, problem-solving, decision making, or investigative activities; gives students the opportunity to work relatively autonomously over extended periods of time; and culminate in realistic products or presentations [31, 48].

Other defining features found in the literature include authentic content, authentic assessment, teacher facilitation but not direction, explicit educational goals [38], cooperative learning, reflection, and incorporation of adult skills [16]. To these features, particular models of PjBL added a number of unique features. Definitions of 'Project-Based Instruction'

include features relating to the use of an authentic (driving) question, a community of inquiry, and the use of cognitive (technology-based) tools [32, 36] and expeditionary learning adds features of comprehensive school improvement, community service, and multidisciplinary themes [47].

The assignment encourages students to use problem-solving skills and collaboration to successfully complete the project. The university will benefit very much in these projects as it involves hardware and/or software that is not available at the institution. The cooperating organization gains by having a low priority project completed without any cost. Based on the experience of the author, companies that do end up implementing the student-developed results are often quite willing to make a contribution (hardware, software or money) to the university [26]. PjBL helps make learning relevant and useful to students by establishing connections to life outside the classroom, addressing larger concerns, enhancing their critical skills, and shaping their learning process by being active participants [7]. What differentiates PjBL from traditional class lessons is that the teacher acts more like a facilitator than like the sage on the stage.

Jazayeri (2004) states that class teaching should integrate with projects. This is often recognized as a very critical issue in Software Engineering education. Replaying the complexity of real-life projects in an educational environment can be impossible. Thus we need to find innovative ways of integrating project work in curricula [19, 30]. They argued that projects should be realistic, but students should be aware of the differences with the real life, in terms of team size, requirements for compatibility with legacy systems and unavailability of real stakeholders.

All the literature reviewed in previous section discussed the significance of skill for employability among undergraduate is an undeniable fact. The undergraduates make the most numbers of contributors for employability rather than post-graduate. The discussion scope is focused on Software Engineering undergraduates in the industry nowadays. Software Engineering is the most significant in the present industry. Due to the failure of product and deliverables in the industry, the literature shows that Requirement Engineering is a factor to that failure. To solve the problem, higher learning education is a platform to assist the Software Engineering undergraduate in enhancing their skills through learning. In addition, the Project-Based Learning should be a method to help the learning environment. It can be supported by the tools. If the students enhance their skill, it will benefit the industry in term of cost in sending the students to training courses and thus employer will attain employability skills. The student will get a skill without wasting their time (to attend other training course) and is more confident on acquiring the job. Software Engineering education; particularly Requirement Engineering education will come out most prominent backed with an effective learning environment in using PjBL to enhance student skills.

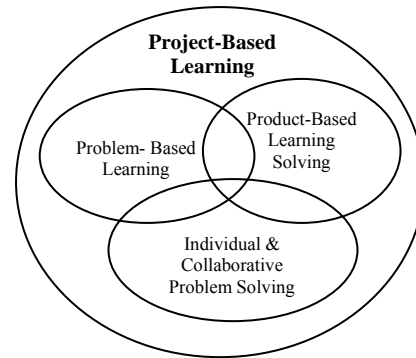


Figure 1. Problem-Based Learning, Individual & Collaborative Problem Solving and Product-Based Learning are a subset to Project-Based Learning

Additionally the researcher identified that Project-Based Learning is a superset for Individual & Collaborative Problem Solving [20, 22, 34], Problem-Based Learning [4, 41, 46] and Product-Based Learning [1, 33]. In contrast, all of them are subset to Project-Based Learning and have an intersection with each other. All major generic skills scheme such as conceptual 'thinking' and interpersonal 'teamwork' skills will be a significant factor in PjBL [6, 44].

If the PjBL wants to be implemented successfully for undergraduate Software Engineering, they need to identify suitable tools, the people with the know-how in PjBL and the needs to come up with a guideline that refers to tertiary culture. By the time the need is identified, the courses have developed, the students trained and the new technology has changed. The education that succeeds will be the one that facilitates lifelong learning, equipping students with the skills they will need to adapt to change. Lacking information on what PjBL practices are most productive, evidence of PjBL's relative effectiveness in comparison to other methods, and an overall framework to guide their planning and collaborations [44].

Moreover, the difficulties in choosing project type based on level (diploma, degree or master level) of project will be faced by educators. The educator should form a project team and identify their roles [11, 47].

As a result, a Project-Based Learning curriculum addresses real-life issues, stresses problem-solving skills, has the teacher serve as facilitator, and lets student's self-assess progress.

## V. CONCLUSION

It can be concluded that current practice of teaching Software Engineering undergraduate in Requirement Engineering is weak. The superset of Project-Based Learning (PjBL) is strongly recommended to enhance Software Engineering undergraduate skills in Requirement Engineering.

Future work will be discussed on the next objective is to design the framework of (PjBL) to be implemented in Requirement Engineering and to measure the effectiveness of Project-Based Learning in teaching Requirement Engineering to enhance Software Engineering undergraduate skills. The researcher will do further research by implementing PjBL in

developing country which is Malaysia as a case study. Furthermore, the supportive tool that is electronic Learning Industrial Environment (eLIN) system will be discussed in next objectives of the research. The eLIN system will be used for student in getting feedback instantly from industry (personnel in charge of the students' projects) in identifying the correctness of student project. The discussion on specification that is suitable to construct the repository/database prototype (eLIN system) that captured and retrieved effectively the changes (major or minor) in Requirement Engineering student's project by industry, educator and students Software Engineering undergraduate (third objective) will be discussed for future work.

- [1] Amsler, A., & Nichols, R., 2008. Blazing a Web CMS Trail at the University of Delaware. Association of Computing Machinery. SIGUCCS'08, October 19–22, 2008, Portland, Oregon, USA, pp:137-140
- [2] Ariffin, R., & Ismail, H. N., 2007. Konsep Keadilan dalam Teori Kecerdasan Pelbagai Menurut Perspektif Islam. Pendidikan Psikologi.
- [3] Bagert, D. J., Port, D. N., & Saideian, H. 2008. Software Engineering Education, Training and Research: The Legacy of Nancy Mead. IEEE, 1: 238-243.
- [4] Bakar, M. S., & Shaikh Ab Rahman, S. N., 2005. A Kick Start in Implementation of PBL in Computer Programming. Proceedings of the 2005 Regional Conference on Engineering Education.
- [5] Baker, A., Navarro, E. O., & Hoek, V. D., 2005. An experimental card game for teaching software engineering processes, Journal of Systems and Software, 75(1): 3-16.
- [6] Bernhart, M., Grechenig, T., Hetzl, J., and Zuser, W., 2006. Dimensions of Software Engineering Course Design. Proceedings of the 28th international conference on Software Engineering Shanghai, China, pp: 667-672.
- [7] Bevan, N., 1999. Quality in use: meeting user needs for quality, Journal of Systems and Software. 49(1), 89-96.
- [8] Boehm, B. (2006). A View of 20th and 21st Century Software Engineering. *ACM*, 12-29.
- [9] Bourque, P., Dupuis, R., Abran, A., Moore, J. W. & Tripp, L., 1999. The Guide to the Software Engineering Body of Knowledge. IEEE Software. 16 (6): pp 35 – 44.
- [10] Carlisle, C., Ibbotson, T., 2005. Introducing Problem-Based Learning into Research Methods Teaching: Student and Facilitator Evaluation, 25(7), 527-541.
- [11] Cheng, B. H., & Atlee, J. M. 2007. Research Directions in Requirement Engineering. Future of Software Engineering, 2007. FOSE '07 In Future of Software Engineering, 2007. pp: 285-303.
- [12] Cheng, B. H., & Atlee, J. M. 2009. Current and Future Research Directions in Requirements Engineering. Design Requirements Engineering: A Ten-Year Perspective, pp. 11-43.
- [13] Clarke, E. (2009). Elizabeth Clarke, 2009. Learning outcomes from business simulation exercises: Challenges for the implementation of learning technologies", Education + Training, 51(5), 448 – 459.
- [14] Crear, J. 2009. CHAOS Summary 2009. The Standish Group International Inc. Boston, Massachusetts: Standish Group's.
- [15] Damian, D., & Chisan, J. 2006. An Empirical Study of the Complex Relationships between Requirements Engineering Processes and Other Processes that Lead to Payoffs in Productivity, Quality, and Risk Management, IEEE Transactions on Software Engineering, 32(7), 433-453.
- [16] Diehl, W., Grobe, T., Lopez, H., & Cabral, C. 1999. Project-based learning: A strategy for teaching and learning. Boston, MA: Center for Youth Development and Education. Corporation for Business, Work, and Learning.
- [17] Duncan, D. (2000). Information Technology Partnership between Industry and Academia. California State University, Hayward. Unpublished.
- [18] Ebmann, B., Hampel, T., & Keil-Slawik, R. 2005. Re-Conceptualizing Learning Environments: A Framework for Pervasive eLearning. Third IEEE International Conference on Pervasive Computing and Communications Workshops. PERCOMW'05, March 08-12, 2005. pp: 322-326.
- [19] Ellis, A. 2001. Student-Centred Collaborative Learning via Face-to-Face and Asynchronous Online Communication: What's the Difference? In G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.), Meeting at the Crossroads. Proceedings of the 18th Annual Conference of the Australian Society for the Computers in Learning in Tertiary Education. pp: 169-178. Melbourne: Biomedical Multimedia Unit, The University of Melbourne.
- [20] Frezza, S., & Cannell, J. 2009. Product-Based Learning in Software Engineering Education. Frontiers in Education Conference, FIE '09. 39th IEEE. 18-21 Oct. 2009. pp: 1-6.
- [21] Gerogiannis, V. & Fitsilis, P. 2006. A Project-Based Learning Approach for Teaching ERP Concepts, International Journal of Learning, Common Ground Publishers, 12(8). pp: 261-268.
- [22] Guzdial, M., Turns, J., Rappin, N., & Carlson, D. (1995). Collaborative support for learning in complex domains. Proceeding CSCL '95 The first international conference on Computer support for collaborative learning. pp: 157-160.
- [23] Haji Abdul Rahman, M. B., Mat Daud, K. A., Abd Ghani, N. A., & Jusoff, K. 2009. Project Based Learning (PjBL) Practices at Politeknik Kota Bharu, Malaysia. Canadian Center of Science and Education. 2(4), p: 140-148.
- [24] Hashim, R., & Mohd Din, M. A. 2008. Implementing Outcome-Based Education Using Project Based Learning at University of Malaya: An Evaluation. Proceedings Paris International Conference on Education, Economy and Society. pp: 51-59.s
- [25] Hayes, J. H., & Biddle, M. 2008. Nancy Mead and Software Engineering Education: Advancements through Action. 21st Conference on Software Engineering Education and Training. pp: 244-249.
- [26] Hayward. 1995. Information technology partnersip, Doris Duncan California State University. 3(2).
- [27] Hoare, C. A. 2006. How did Software get so Reliable without Proof? 1-17.
- [28] IBM. (2010). Collaborative Application Lifecycle Management Solutions from IBM Rational. International Business Machines Corporation.
- [29] Jackson, D. E. 2007. Software for Dependable Systems-Sufficient Evidence? IEEE.
- [30] Jayazeri, M. 2004. Education of a Software Engineer. IEEE.
- [31] Jones, B., Rasmussen, C. M., & Moffitt, M. 1997. Real-life problem solving: A collaborative approach to interdisciplinary learning. Washington: American Psychological Association.
- [32] Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. 1994. A collaborative model for helping middle-grade science teachers learn project-based instruction. 94, 483-497.
- [33] Leifer, L. (1995). Evaluating Product-Based-Learning Education. 28.
- [34] Marshall, S. P., & al, e. (2003). In the Service of Learning, Getting to the heart of Problem-based Learning. Summary Report, PBL Symposium.

AUTHORS PROFILE

- [35] Martin, R., Villeneuve-Smith F., Marshall, L., & McKenzie, E., (2008). Employability skills explored research report. Learning and Skills Network.
- [36] Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Blunk, M., Crawford, B., Kelley, B., et al. 1994. Enacting project-based science: Experiences of four middle grade teachers. 94, 517-538.
- [37] Mead, N. R. 2009. Software Engineering Education: How far We've Come and How Far We Have To Go. 82 571-575.
- [38] Moursund, D. 1999. Project-Based Learning using Information Technology. Eugene or International Society for Technology in Education.
- [39] Naur, P., & Randell, B. 1969. Software Engineering Report of a conference sponsored by the NATO Science Committee Garmisch Germany 7th-11th October 1968 . Germany: Scientific Affairs Division, NATO.
- [40] Parnas, D. L. 1999. Software Engineering Programs Are Not Computer Science Programs. IEEE, 16 (6), 19-30.
- [41] Ragan, E. D., Frezza, S., & Cannell, J. (2009). Product-Based Learning in Software Engineering Education.
- [42] Rashid, R. A., Abdullah, R., Zaharim, A., Ghulman, H. A. & Masodi, M. S., 2008. Engineering Students Performance Evaluation of Generic Skills Measurement: ESPEGS Model. 5th WSEAS / IASME International Conference on Engineering Education.
- [43] Rugarcia, A. Felder, R. M., Woods, D. R. & Stice J. E., 2000. The Future of Engineering Education I. A vision for a New Century. Chem. Engr. Education , 34 (1), 16-25.
- [44] Rusu, Ad., Rusu, Am., Docimo, R., Santiago, C. & Paglione, M., 2009. Academia-academia-Industry Collaborations on Software Engineering Projects Using Local-Remote Teams. ACM , 301-305.
- [45] Sarif, S. M., & Shiratuddin, N. (2009). md-Matrix: An assistive learning tool in blended project-based learning for mobile development course.
- [46] Savery, J. R. (2006). Overview of Problem-based Learning: Definitions and Distinctions The Interdisciplinary Journal of Problem-based Learning. 1(1).
- [47] Thomas, J. W. 2000. A Review of Research on Project-Based Learning.
- [48] Thomas, J. W., Mergendoller, J. R., & Michaelson, A. 1999. Project-based learning: A handbook for middle and high school teachers.
- [49] Turhan, B., & Bener, A. 2007. A Template for Real World Team Projects for Highly Populated Software Engineering Classes. Software Engineering, 2007. ICSE 2007. 29th International Conference on IEEE Computer Society , 748-753.
- [50] Windbladh, K. (2004). Requirement Engineering Closing the Gap Between Academic Supply and Industry Demand. The ACM Student Magazine. 10 (4), August 2004.
- [51] Wolff, S. J. 2002. Design Features for Project-Based Learning.
- [52] Zave, P. (1997). Classification of Research Efforts in Requirements Engineering. Journal ACM Computing Surveys (CSUR). 29 (4), 315-321.



**Nor Azlana Akmal Jamaludin** is a Lecturer for Bachelor of Science Software Engineering and Head of Developer for Master Degree Software Engineering program at Universiti Industri Selangor. She received the Master Degree in Computer Science (Real-Time Software Engineering) from Advanced Informatics School (formerly known as Centre for Advanced Software Engineering (CASE), Universiti Teknologi Malaysia, in 2004. Currently, she is doing her Doctorate in Computer Science, specialize in Software Engineering at UTM and supervise by Prof. Dr. Shamsul Sahibuddin. She is a member of Malaysian Software Engineering Interest Group, Malaysia. Her field of expertise is in software requirement, analysis, system integration, and e-learning. Her current research interest is on techniques that can enhance skill among Software Engineering undergraduate of higher institutions in Malaysia using machine learning system.



**Shamsul Sahibuddin** received the PhD in Computer Science at Aston University, Birmingham, United Kingdom in 1998 and Master Science (Computer Science) at Central Michigan University, Mt. Pleasant, Michigan in 1988. He is currently a Dean / Professor at Advanced Informatics School, Universiti Teknologi Malaysia (UTM) and has been a Member of the Program Committee for Asia-Pacific Conference on Software Engineering. His field of expertise is in software requirement, software modelling technique, software development and software process improvement. He has been very active in scholarly journals-book writing and publishing.



**Hj. Kamaruzaman Jusoff** was born in Kota Bharu, Kelantan, Malaysia, in 1958. He received the PhD in Forest Engineering from Cranfield University, England, UK in 1992. He is currently a Professor (Grade B) with Universiti Putra Malaysia and has been a Visiting Professor with The Harvard Forest, Harvard University, Kyoto's Graduate School of Engineering, Yale's Center for Earth Observation and an Erasmus Mundus Professor in Lund University, Sweden and ITC, The Netherlands. His field of expertise is in airborne hyperspectral image processing. Research and higher education leadership especially in creating succession planning models and talent pools have also being his current field of interests. Hj. Kamaruzaman has been very active in scholarly journals-book writing and publishing and has published 7 books and more than 250 citation index/impact factor journal papers. Administratively, used to hold several posts in UPM as Deputy Dean, Faculty of Forestry, UPM Kampus Bintulu Designate Director, MACRES Deputy Director (R&D) while being active doing some consultancies on airborne hyperspectral sensing for forestry, agriculture, marine and Search-and-Rescue operations.



**Nur Hafizah Hidayat** is a Lecturer of Information Systems at the University Industrial Selangor (UNISEL), Malaysia. She also serves as Head of Information Technology in the Center for Research and Innovation for the UNISEL. She obtained her Bachelor Science in Information Technology from Universiti Teknologi Mara, Malaysia and Master Science in Information Management from Leeds Metropolitan University, United Kingdom. Her research interest includes information system management, project management, business intelligence and open source system.

# Adaptation of GQM Method for Evaluating the Performance of Software Project Manager

Nagy Ramadan Darwish

Department of Computer and Information Sciences,  
Institute of Statistical Studies and Research,  
Cairo University  
Cairo, Egypt  
drnagyd@yahoo.com

Mahmoud M. Abd Ellatif

Department of Information Systems,  
Faculty of Computers and Information,  
Mansoura University  
Mansoura, Egypt  
drmmlatif@yahoo.com

**Abstract**—This paper is concerned with evaluating the performance of software project manager using Goal Question Metrics (GQM) method. It clarifies the Software Project Management (SPM) domains and the performance metrics of each domain. This paper presents the basic concepts of GQM method. Based on a combination of statistical techniques, this paper presents how to apply GQM method to evaluate the performance of a software project manager. A software company can use the proposed approach to track, evaluate, control, correct, and enhance the performance of software project managers to increase the success rate of software projects.

**Keywords** *Software ;Project Manager; Performance; Evaluation - GQM – Metrics – Performance Report*

## I. INTRODUCTION AND PROBLEM DEFINITION

SPM is the on-going activities for planning, organizing, directing, and controlling progress to develop an acceptable system, i.e. conform to the quality standards within the allocated time and budget [11]. The mismanaged projects may lead to: unfulfilled or unidentified requirements, uncontrolled change of project scope, uncontrolled change of technology, uncontrolled risk of the project, uncontrolled subcontracting and integration, cost overruns, and/or late delivery [12]. The failure rate of large software projects is larger than the success rate [16]. Therefore, there is a need to track, evaluate, control, correct, and enhance the performance of software project managers to increase the success rate of software projects.

GQM method can be used to evaluate the performance of software project managers. This method was developed for multi-purpose evaluation of software. GQM method consists of three steps: determination of a Goal, construction of sets of Questions which have possible answers. The last step is analytic of the set of metrics, which consists of weight coefficient for each set of answers [22].

## II. SPM DOMAINS AND PERFORMANCE METRICS

SPM activities can be organized in nine domains [9]: integration management, scope management, schedule management, costs management, quality management, human resources management, communications management, risk management, and procurement management. Each domain includes a set of activities related to a specific field in SPM practices.

For example, scope management includes the following activities:

- Identifying the project background.
- Assessing the initial feasibility of the project.
- Defining project scope and deliverables.
- Listing project assumptions and constraints.
- Verifying the project scope.
- Establishing procedures for tracking project progress.
- Assessing the project feasibility.
- Managing project scope changes.
- Tracking project progress.

System improvement requires measurement and analysis [21]. Performance measurements are used in project management and quality processes to determine and communicate status and accomplishments measured against specific objectives, schedules, and milestones. These measurements extend to include delivery of desired products and services to customers, whether external or internal [3]. Performance measurement can be useful to improve future work estimates [15]. Performance measurement is the ongoing monitoring and reporting of project accomplishments, particularly progress towards pre-established goals. Performance measures may address: the type or level of project activities conducted, the direct products and services delivered by a program, and/or the results of those products and services [2].

A metric is a quantitative measure of the degree to which a system, component, or process possesses a given attribute [7]. Performance metrics should be objective, timely, simple, accurate, useful, and cost-effective. The performance metrics can be divided into three basic categories [2]: measures of efforts, measures of accomplishments, and measures that relate efforts to accomplishments.

The researchers elaborated a set of performance metrics for scope management. These elaborated performance metrics can be calculated at specific time check points such as project milestones. In addition, they can be calculated for a given time interval such as a week, month, and so on. Therefore, they can be calculated weekly as an integral part of project progress report. The elaborated performance metrics for scope management include:



- Percentage of users involved in defining scope and deliverables vs. planned.
- Percentage of project deliverables achieved vs. planned.
- Percentage of project deliverables reviewed and approved vs. achieved.
- Percentage of major milestones met vs. planned.
- Percentage of project team meetings achieved vs. planned.
- Percentage of scope change requests subjected to feasibility studies vs. all requests.
- Percentage of scope change requests subjected to integration tests vs. all requests.
- Percentage of scope change requests subjected to configuration management tests vs. all requests.
- Percentage of preliminary feasibility studies achieved vs. required at the initiation phase.
- Return On Investment (ROI) calculated for the potential preliminary project's costs at the initiation phase.
- ROI calculated for the potential detailed project's costs at the planning phase.
- Payback period calculated for the potential preliminary project's costs at the initiation phase.
- Payback period calculated for the potential detailed project's costs at the planning phase.
- Percentage of detailed feasibility studies achieved vs. required at the planning phase.

### III. GOAL QUESTION METRICS METHOD

Victor Basili and et al at Maryland University developed a goal oriented approach for measurement [5]. This approach depends on three steps:

- Set goals specific to needs in terms of purpose, perspective, and environment.
- Refine the goals into quantifiable questions that are tractable.
- Deduce the metrics and data to be collected (and the means for collecting them) to answer the questions.

In GQM method, each goal generates a set of quantifiable questions that attempt to define and quantify this goal. The question can only be answered relative to, and as completely as, the available metrics allow. In GQM, the same question can be used to define multiple goals. Also, metrics can be used to answer more than one question. Unfortunately, this approach suits to mature and well-understood problem areas.

### IV. THE PROPOSED APPROACH FOR EVALUATING PERFORMANCE OF SOFTWARE PROJECT MANAGER

GQM method can be used to evaluate the performance of software project managers. Each SPM domain has a goal, for example schedule management aims to administrate and control of the finite resource of time. The three conventional measures of project success are budget, schedule, and functionality. The project manager must manage the schedule carefully for preventing or correcting any slippages. Each goal is decomposed into several questions, and each question can be answered by a metric or more. Figure (1) illustrates the

adaptation of GQM method to fit the performance evaluation of software project manager. Each software company includes a quality group to evaluate and track the performance of the software project managers.

Calculating metrics is a simple process because it depends on simple or known statistical or mathematical formulas such as Return On Investment (ROI), payback, percentage, ratio, cost deviation, and time deviation (in hours, days, weeks, or months).

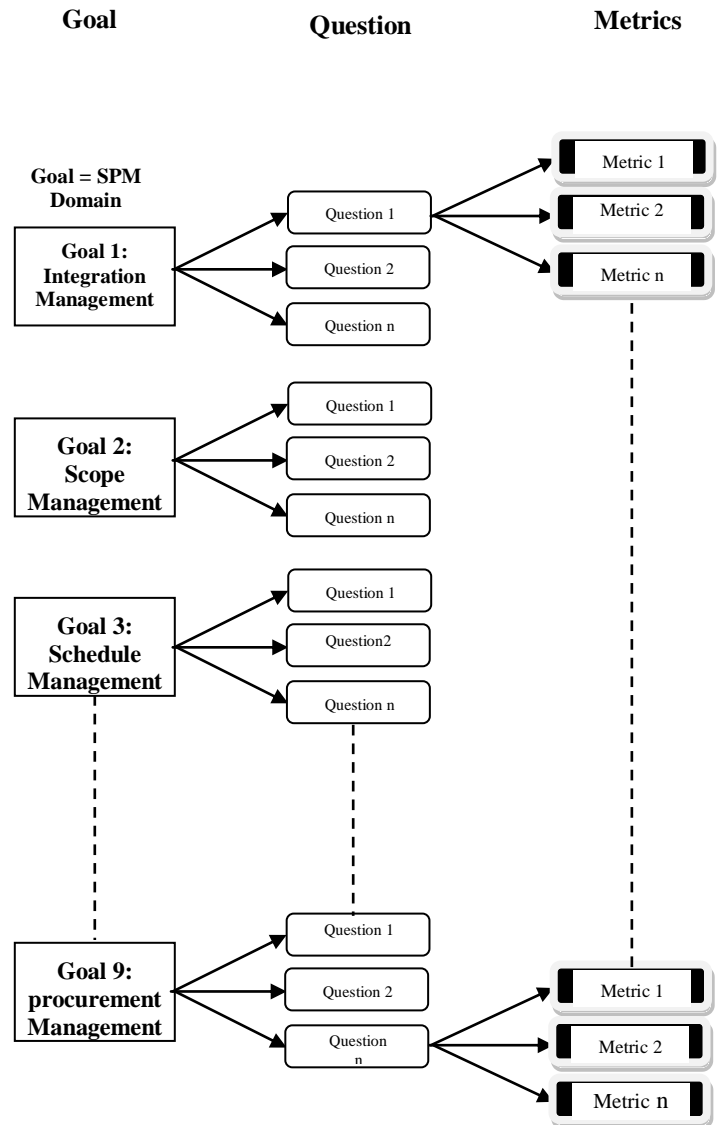


Figure (1): GQM Method and SPM domains.

Figure (2) illustrates a general flowchart that presents the proposed approach. The proposed approach includes the following main procedures:

- Define the goals, questions, and performance metrics.
- Calculate the value of each performance metric and compare it with the accepted range of the metric value.
- Interpret and analyze the performance report.

#### A. Define the Goals, Questions, and Performance Metrics

Each SPM domain have a goal, therefore there are nine goals: integration management, scope management, schedule management, costs management, quality management, human resources management, communications management, risk management, and procurement management. Each goal is decomposed into several questions, and each question can be answered by a metric or more. The definition of each metric should include the mathematical or statistical techniques for calculating this metric.

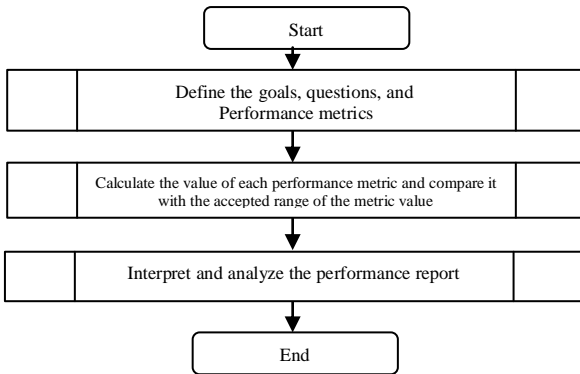


Figure (2): The proposed approach for Evaluating Performance of Software Project Manager.

In addition, the definition of each metric should include the accepted range of the metric value. The time check points for calculating performance metrics should be determined. These metrics can be calculated weekly as a part of project progress report. The project manager should be involved in this process. The quality group should present the SPM performance metrics to the project manager and deal with his objections by clarifying, negotiating, or modifying these metrics. Previous experience from similar projects can be useful in this process. In addition, this process can be achieved with the assistance of external consultants to define and validate the SPM performance metrics. Figure (3) illustrates the steps of this procedure.

1. Select a goal from the list of goals of SPM domains (nine goals).
2. Define a list of questions related to the selected goal.
3. Select a question from the list of questions.
4. Define a list of performance metrics related to the selected question.
5. Select a performance metric to be defined in detail.
6. Define the mathematical or statistical technique for calculating the selected metric.
7. Define time check points for calculating the selected metric.
8. Define the accepted range of the metric value of the selected metric.
9. Negotiate the project manager and deal with his objections by clarifying or modifying the performance metric.
10. Check the list of performance metrics. If it is not empty, then go to step 5.
11. Check the list of questions. If it is not empty, go to step 3.
12. Check the list of goals. If it is not empty, go to step 1.

Figure (3): The definition of the Goals, Questions, and Performance Metrics.

Table (1) illustrates an elaborated list of questions and performance metrics for scope management.

Question	Metric
Q1: Is the users involved in defining scope and deliverables?	1. Percentage of users involved in defining scope and deliverables vs. planned.
Q2: Is the project deliverables planned, achieved, reviewed, and approved.	1. Percentage of project deliverables achieved vs. planned. 2. Percentage of project deliverables reviewed and approved vs. achieved.
Q3: Is the major milestones of the project met?	1. Percentage of major milestones met vs. planned.
Q4: Are the project team meetings achieved?	2. Percentage of project team meetings achieved vs. planned.
Q5: Are the scope change requests dealt well?	1. Percentage of scope change requests subjected to feasibility studies vs. all requests. 2. Percentage of scope change requests subjected to integration tests vs. all requests. 3. Percentage of scope change requests subjected to configuration management tests vs. all requests.
Q6: Are the feasibility studies do?	1. Percentage of preliminary feasibility studies achieved vs. required at the initiation phase. 2. Return On Investment (ROI) calculated for the potential preliminary project's costs at the initiation phase. 3. ROI calculated for the potential detailed project's costs at the planning phase. 4. Payback period calculated for the potential preliminary project's costs at the initiation phase. 5. Payback period calculated for the potential detailed project's costs at the planning phase. 6. Percentage of detailed feasibility studies achieved vs. required at the planning phase.

Table (1): The elaborated list of questions and performance metrics for scope management.

#### B. Calculate the Performance Metrics and compare it with the accepted range

The second procedure of the proposed approach is calculating the performance metrics for the questions of a specific goal. Calculating performance metrics is a simple process because it depends on simple or known statistical or mathematical formulas. So, the researchers don't focus on calculating performance metrics. After calculating the value of the performance metric, the quality groups compare this value with the accepted range of the metric value. If the metric value is out of the accepted range, the quality groups add a deviation note to the performance report. The quality groups prepare a performance report that must include the metric value, accepted range, and any deviation notes. Some performance metrics may be Not Applicable (NA) in some specific cases [19]. So, during computing the value of the performance metrics, the NA

metrics are eliminated. Figure (4) illustrates the steps of this procedure.

### C. Interpret and Analyze the Performance Report

The third procedure of the proposed approach is interpreting and analyzing the performance report. The quality group should report their interpretation to their top management. If the performance deviations is not accepted, top management may take corrective actions or inform the project manager to take corrective actions. The value of performance deviations should be analyzed to discover the weaknesses and strengths of project management practices. This analysis can be used to reduce or avoid many risks or obstacles that may be encountered in later points of time or in next software projects.

## V. CONCLUSION

The main objective of this paper was proposing an approach for evaluating the performance of software project manager using GQM method. In GQM method, each goal generates a set of quantifiable questions that attempt to define and quantify this goal. The question can only be answered relative to, and as completely as, the available metrics allow. Evaluating the performance of software project manager is helpful for increasing capability level and productivity, improving quality, tracking project progress, and assessing project status.

We conclude that the roles of quality group are very important in software projects. They can use the list of performance metrics and the proposed approach to evaluate and track the performance of the software project manager. In addition, we conclude that special emphasis must be given to performance metrics in software projects to discover and avoid the weaknesses of practices. On the other hand, the strengths of project management practices must be utilized and encouraged.

## VI. FUTURE WORK

For future work, the following points are expected to be focused:

- Improvements in integration and scope management of software projects.
- Improvements in schedule and cost management of software projects.
- Achieving higher levels in Capability Maturity Model Integration (CMMI) for IT companies.
- Enhancements in the quality of e-government projects.

## REFERENCES

- [1] Chang E. Koh, Victor R. Prybutok, and Xiaoni Zhang, "Measuring e-government readiness", *Journal of Information & Management* 45 pages: 540–546, 2008.
- [2] Department of Energy (DOE), "Basic Performance Measures for Information Technology Projects", 2002, <http://cio.doe.gov/ITReform/sqse/download/PE-WI-V3-011502.doc>
- [3] Department of Energy (DOE), "IT Project Management Review Process Guide", 2003, <http://cio.doe.gov/ITReform/sqse/download/QR-CP-F3-011403.doc>
- [4] El Hachemi Alikacem and Houari A. Sahraoui, "Rule-Based System for Flaw Specification and Detection in Object-Oriented Programs", 2010, [www.crim.ca/Publications/2010/documents/plein\\_texte/ASD\\_AliE\\_al\\_QAOOSE\\_2010.pdf](http://www.crim.ca/Publications/2010/documents/plein_texte/ASD_AliE_al_QAOOSE_2010.pdf)
- [5] Fenton, Norman, Robin Whitty and Yoshinori Iizuka, "Software Quality Assurance and Measurement", International Thomson Computer Press, 1995.
- [6] Girish H. Subramanian, James J. Jiang, and Gary Klein, "Software quality and IS project performance improvements from software development process maturity and IS implementation strategies", the *Journal of Systems and Software* 80, 2007, pages: 616–627
- [7] IEEE Standard 610, "Glossary of Software Engineering", 1990.
- [8] Ince, "Software Quality Assurance - A Student Introduction", McGraw-hill international (UK) limited, 1995.
- [9] Information Systems Audit and Control Foundation (ISACF), "Project Management: Skill and Knowledge Requirements in an Information Technology Environment", 2002.
- [10] ISO 10006, "Quality Management Systems – Guidelines for Quality management in Projects", Second Edition, 2003.
- [11] Jeffrey A. Hoffer, Joey F. George, and Joseph S. Valacich, "Modern System Analysis and Design", Addison Wesley Longman, Inc, 1999.
- [12] Jeffrey L. Whitten, Lonnie D. Bentley and Kevin C. Dittman, "System Analysis and Design Methods", Fifth edition, Mc Graw Hill Companies, Inc, 2001.
- [13] Joel Henry, "Software Project Management – A Real-World Guide to Success", Pearson Education, Inc, 2004.
- [14] Jolyon E. Hallows, "Information Systems Project Management: How to Deliver Function and Value in Information Technology Projects", AMACOM, a division of American Management Association, 1998.
- [15] Karl E. Wieggers, "A Software Metrics Primer", Process Impact, 1999, [http://www.processimpact.com/articles/metrics\\_primer.html](http://www.processimpact.com/articles/metrics_primer.html)
- [16] Paul Dorsey, "Top 10 Reasons Why Systems Projects Fail", 2000, <http://www.duclican.com/papers/top%2010%20reasons%20why%20systems%20projects%20fail.htm>
- [17] Paul Goodman, "Software Metrics: Best Practices for Successful IT Management", Rothstein Associates, 2004.
- [18] Robert T. Futrell, Donald F. Shafer, and Linda I. Safer, "Quality Software Project Management", Prentice Hall PTR, 2002.
- [19] William E. Perry, "Quality Assurance for Information Systems: Methods, Tools, and Techniques", QED technical publishing Group, 1991.
- [20] Yael Dubinsky and Orit Hazzan, "Using a role scheme to derive software project metrics", *Journal of Systems Architecture* 52 pages: 693–699, 2006.
- [21] Zahedi, Fatemeh, "Quality Information Systems", Boyd & Fraser Publishing Company - a Division of International Thomson Publishing Inc., 1995.
- [22] Zdena Dobešová1 and Dagmar Kusendová2, "Goal-Question-Metric method for evaluation of cartographic functionality in GIS software", proceedings GIS Ostrava, 2009, [www.geoinformatics.upol.cz/app/visehrad/images/GISOstrava.pdf](http://www.geoinformatics.upol.cz/app/visehrad/images/GISOstrava.pdf)

# Adaptive E-Learning System based on Semantic Web and Fuzzy Clustering

*Khaled M. Fouad*

Computer Science Dep. - Community College - Taif University,  
Kingdom of Saudi Arabia (KSA);  
Central Laboratory of Agriculture Expert Systems (CLAES), Egypt  
kmfi25@yahoo.com

*Shehab Gamalel-Din*

Computers Dept, Faculty of Science, King Abdul-Aziz  
University, Kingdom of Saudi Arabia (KSA)  
smostafa@kau.edu.sa

*Mofreh A. Hogo*

Computer Science Dep. – Computers and Information Systems  
College - Taif University, Kingdom of Saudi Arabia (KSA).  
mofreh\_hogo@yahoo.com

*Nagdy M. Nagdy*

Systems and Computers Engineering Dep. – Faculty of  
Engineering – AlAzhar University, Egypt.  
prof\_nagdy@yahoo.com

**Abstract—** This work aims at developing an adaptive e-learning system with high performance to reduce the challenges faces e-learners, the instructors and provides a good monitoring system for the complete e-learning systems as well as the system structure. The work presents the different phases for the system development of the adaptive system as: the first stage is the collection of the e-learners documents, the second stag is the documents representation including the frequency count and the weighting of the documents with its frequencies, the third stage is the prediction and clustering of e-learners interests using the fuzzy clustering method and the statistical K-means clustering method. The results obtained from this work shows that we have to have different e-learners ontologies using the results of the clustering methods which reflect the e-learners interests. Finally the work concluded the suggestions as well as the recommendations for the instructors and the systems administrators.

**Keywords-component;** E-Learning; Semantic Web; Fuzzy Ckustering; User model; User Model Representation.

## I. INTRODUCTION

Electronic learning or E-Learning [1] is interactive learning in which the learning content is available online and provides automatic feedback to the student's learning activities, it is much like computer-based training (CBT) and computer-aided instruction (CAT), but the point is that it requires Internet for access to learning material and for monitoring the student's activities. E-Learners usually can communicate with their tutors through the Internet.

Semantic web could offer unprecedented support to the network teaching in semantic query, meaning construction, knowledge acquisition and sharing and collaborative learning. Simultaneously semantic web also provides the support to describe semantics of learner characteristic in the learner model, and makes it possible to share learner model between systems. So we need to construct the learner model and clustering it; to simplify contents search that is based on that learner profile, in adaptive learning system based on semantic web.

## A. ADAPTIVE E-LEARNING

In the context of e-learning [2], adaptive e-learning systems are more specialized and focus on the adaptation of learning content and the presentation of this content. According to [3], an adaptive system focuses on how the knowledge is learned by the student and pays attention to learning activities, cognitive structures and the context of the learning material.

The structure of an adaptive e-learning system is shown in Fig. 1.

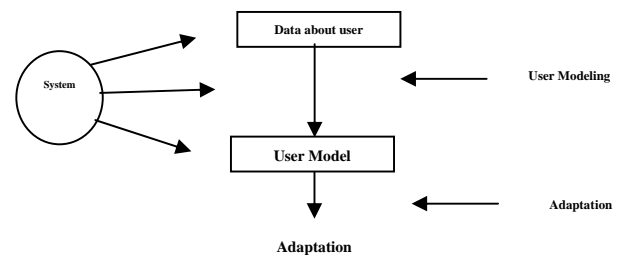


Fig. 1: The Structure of an Adaptive System [2]

The system intervenes at three stages during the process of adaptation. It controls the process of collecting data about the learner, the process of building up the learner model (learner modeling) and during the adaptation process. It is not feasible in conventional WBE to create static learning material [1] that can be read in any arbitrary sequence, because of many interdependences and prerequisite relationships between the course pages. However, adaptive hypermedia (AH) methods and techniques make it possible to inform learners that certain links lead to material they are not ready for, to suggest visiting pages the learner should consult, or automatically provide additional explanations at the pages the learner visits, in order to scaffold his/her progress. Adaptive educational hypermedia systems (AEHSs) apply different forms of learner models to adapt the content and the links of hypermedia course pages to the learner. AEHSs support adaptive learning, using technology to constantly measure the learner's knowledge and progress in order to adapt learning content delivery, presentation, feedback, assessment, or environment to the learner's needs, pace, preferences, and goals. Such systems make predictions of what the learner needs to attain his/her goals, respond to such needs,

allocate resources, implement change, and thus improve personalization of the learning process. The system can be designed to use predictive strategies prior to instruction delivery and learning sessions, during the instruction (based on the learner's interaction), or both.

The open adaptive learning environment [4] is in which learners dynamically select a learning route suitable to their needs and profile. The proposed environment is based on the IEEE/IMS learning object metadata (LOM) standard. The nature of adaptations provided by this environment are centered on the learner, and allow the LO to adapt to the evolving learner's model in terms of background, learning modalities, and learning styles. It was summarized [1] the role of personalization in learning environments as follows:

- Personalized learning environments enable one-to-one or many-to-one learning paradigms (one teacher - one learner, and many teachers – one learner), contrary to traditional learning environments that always adopt one-to-many learning paradigm (one teacher, many students);
- Personalized learning environments impose no constraints in terms of learning time, location, etc., whereas traditional ones are fairly restricted by the learning setting;
- Personalized learning environments recognize the huge variety in the learner's characteristics and preferences in terms of the learning style, media, interests, and the like, and adapt instruction according to them; traditional ones are usually designed for the "average learner";
- Personalized learning environments tailor instruction to suit the learner's requirements (self-directed learning); in traditional learning environments, the curriculum, learning units, and the selection and sequencing of learning material are determined by the tutor.

## B. E-LEARNING AND SEMANTIC WEB

It is generally agreed that credible evidence for mastery of learned material [5] is the goal of instructors. While educators and domain experts agree that decoding meaning from text plays a critical role in the acquisition of knowledge across all disciplines, what particular evidence of mastery is required and what lends credibility to such evidence are the subjects of a lively debate among experts in the learning community. The need for new methods for semantic analysis of digital text is now widely recognized in the face of the rising tide of information on the Web. The layered model [6] for the Semantic Web as shown in Fig. 2 puts the relationship among ontology description languages, RDF and RDF Schema, and XML in a better perspective. The bottom layer offers character encoding (Unicode) and referencing (URI) mechanisms. The second layer introduces XML as the document exchange standard.

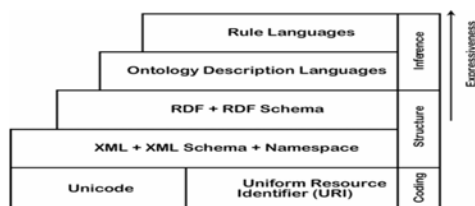


Fig. 2: The Architecture of the Semantic Web

The third layer accommodates RDF and RDF Schema as mechanisms to describe the resources available on the Web. As such, they may be classified as lightweight ontology languages. Full ontology description languages appear in the fourth layer as a way to capture more semantics. The topmost layer introduces expressive rule languages. The semantic web [7] is a space understandable and navigable by both human and software agents. It adds structured meaning and organization to the navigational data of the current web, based on formalized ontologies and controlled vocabularies with semantic links to each other. From the E-Learning perspective, it aids learners in locating, accessing, querying, processing, and assessing learning resources across a distributed heterogeneous network; it also aids instructors in creating, locating, using, reusing, sharing and exchanging learning objects (data and components). The semantic web-based educational systems need to interoperate, collaborate and exchange content or re-use functionality. A key to enabling the interoperability is to capitalize on (1) semantic conceptualization and ontologies, (2) common standardized communication syntax, and (3) large-scale service-based integration of educational content and functionality provision and usage. The vision of the semantic web-based E-Learning is founded on the following major premises:

- Machine-understandable educational content
- Shareable educational ontologies, including:
  - Subject matter ontologies
  - Authoring ontologies (modeling authors' activities)
- Educational semantic web services, for supporting:
  - Learning, e.g., information retrieval, summarization, interpretation (sense-making), structure-visualization, argumentation, etc.
  - Assessment, e.g., tests and performance tracking
  - Collaboration, e.g., group formation, peer help, etc.
- Semantic interoperability

Semantic interoperability, the key promise of the semantic web, is defined as a study of bridging differences between information systems on two levels as following:

- on an access level, where system and organizational boundaries have to be crossed by creating standardized interfaces that share system-internal services in a loosely-coupled way; and
- On a meaning level, where agreements about transported data have to be made in order to permit their correct interpretation. Interoperability requires the use of standard SW languages for representing ontologies, educational content, and services.

## C. THE ONTOLOGIES

The word ontology comes from the Greek ontos, for being, and logos, for word. In philosophy, it refers to the subject of existence, i.e. to the study of being as such. More precisely, it is the study of the categories of things that exist or may exist in some domain [1]. Domain ontology explains the types of things in that domain. Ontology [8] comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic

for some particular topic. Ontologies applied to the Web are creating the Semantic Web. Ontologies provide the necessary armature around which knowledge bases should be built, and set grounds for developing reusable Web-contents, Web-services, and applications. Ontologies facilitate knowledge sharing and reuse, i.e. a common understanding of various contents that reaches across people and applications. Technically, an ontology is a text-based piece of reference-knowledge, put somewhere on the Web for agents to consult it when necessary, and represented using the syntax of an ontology representation language. There are several such languages around for representing ontologies, for an overview and comparison of them. It is important to understand that most of them are built on top of XML and RDF.

The most popular higher-level ontology representation languages were OIL (Ontology Inference Layer) and DAML+OIL. An ontology developed in any such language is usually converted into an RDF/XML-like form and can be partially parsed even by common RDF/XML parsers. Of course, language-specific parsers are necessary for full-scale parsing. There is a methodology for converting an ontology developed in a higher-level language into RDF or RDFS. In early 2004, W3C has officially released OWL (Web Ontology Language) as W3C Recommendation for representing ontologies. OWL is developed starting from description logic and DAML+OIL. The increasing popularity of OWL might lead to its widest adoption as the standard ontology representation language on the Semantic Web in the future. Essentially, OWL is a set of XML elements and attributes, with well-defined meaning, that are used to define terms and their relationships (e.g., Class, equivalent Property, intersection Of, union Of, etc.).

#### D. LEARNER MODEL AND PROFILE

The behavior of an adaptive system [9] varies according to the data from the learner model and the learner profile. Without knowing anything about the learner, a system would perform in exactly the same way for all learners. It was described the application of learner models as follows:

An extensive learner model must contain information about the learner's domain knowledge, the learner's progress, preferences, goals, interests and other information about the learner, which is important for the used systems. Learner models can be classified according to the nature and form of information contained in the models. Considering the subject domain, the information stored in a learner model can be divided into two major groups: domain specific information and domain independent information.

We examined two of the most important and well-developed standards - the PAPI standard [10] and the IMS LIP standard [11]. Both standards deal with several categories for information about a learner. These standards have been developed from different points of view. The PAPI standard reflects ideas from intelligent tutoring systems where the performance information is considered as the most important information about a learner. The PAPI standard also stresses on the importance of inter-personal relationships. On the other

hand the LIP standard is based on the classical notion of a CV and inter-personal relationships are not considered at all.

#### E. K-MEANS CLUSTERING METHOD

Clustering of objects is as ancient as the human need for describing the salient characteristics of men and objects and identifying them with a type. Therefore, it embraces various scientific disciplines: from mathematics and statistics to biology and genetics, each of which uses different terms to describe the topologies formed using this analysis.

The simplest and most commonly used algorithm, employing a squared error criterion is the K-means algorithm [23]. This algorithm partitions the data into K clusters ( $C_1, C_2, \dots, C_K$ ), represented by their centers or means. The center of each cluster is calculated as the mean of all the instances belonging to that cluster. The algorithm [23] starts with an initial set of cluster centers, chosen at random or according to some heuristic procedure. In each iteration, each instance is assigned to its nearest cluster center according to the Euclidean distance between the two. Then the cluster centers are re-calculated. The center of each cluster is calculated as the mean of all the instances belonging to that cluster:



Where  $N_k$  is the number of instances belonging to cluster k and  $\mu_k$  is the mean of the cluster k.

A number of convergence conditions are possible. For example, the search may stop when the partitioning error is not reduced by the relocation of the centers. This indicates that the present partition is locally optimal. Other stopping criteria can be used also such as exceeding a pre-defined number of iterations. Figure 3 presents the pseudo-code [23] of the K-means algorithm.

**Input:**  $S$  (instance set),  $K$  (number of cluster)

**Output:** clusters

- 1: Initialize K cluster centers.
- 2: while termination condition is not satisfied do
- 3: Assign instances to the closest cluster center.
- 4: Update cluster centers based on the assignment.
- 5: end while

Fig. 3. K-means Algorithm.

The rest of the paper is organized as following: Section 2 is reserved for the related works; section 3 introduces the proposed system structure architecture, section 4 presents the experiments design and results analysis, section 5 presents the concluded suggestions and recommendations to improve the system performance, finally section 6 concludes the work and introduces the future work.

## II. RELATED WORKS

An accurate representation of a learners interests [12], generally stored in some form of learner model, is crucial to the performance of personalized search or browsing agents. Learner model is often represented by keyword/concept vectors or concept hierarchy. The acquired model can then be



used for analyzing and predicting the future learner access behavior. Learner model may be built explicitly, by asking learners questions, or implicitly, by observing their activity.

In [13], authors investigated the techniques to create a user profile automatically using the ontological approach. They used a framework to gather the user information from different search space where user's details could be found. The details include user's general information to specific preferences. They used Meta search in user's blog, personal/organization web page, and any other sites to collect information about user. This information is assigned to a pre-structure hierarchy or in reference ontology to create an initial user profile. More clearly, initial profile is learned by the concept/document collected from user's details. In traditional user profiling system feature extraction from document is done by vector space model or considering term frequency, tf-idf methods only. In this research, authors considered WordNet and Lexico-Syntactic pattern for hyponyms to extract feature from document. This profile further improved by taking collaborative user methods. Where, they found a group of users with similar interest by taking similarity score among them. After that an ontology matching approach is applied to learn the profile with other similar user which is called improved profile. [3] Introduced a method for learning and updating a user profile automatically. The proposed method belongs to implicit techniques – it processes and analyzes behavioral patterns of user activities on the web, and modifies a user profile based on extracted information from user's weblogs. The method relies on analysis of web-logs for discovering concepts and items representing user's current and new interests. Those found concepts and items are compared with items from a user profile, and the most relevant ones are added to this profile. The mechanism used for identifying relevant items is built based on a newly introduced concept of ontology-based semantic similarity. There are many different methods to Construct Learner Models [2]: Machine Learning Methods, Bayesian Methods, Overlay methods, Stereotype methods, Plan Recognition. Update of Learner Models methods are: Analysis of Learner Responses, Analysis of the Process of Problem Solution, Analysis of Learner Actions, Discounting Old Data.

In [14], student model mainly included the cognitive model and the interest model. Cognitive model mainly pay attention to learner background knowledge, study style and cognition level. Through synthesizing the domestic and foreign research practice authors proposed to use Solomon Study Style Measure Meter as preceding measure to test learner's study style. Regarding the student's cognition level's estimate they took the thought of fuzzy set. It was combined the ontology and concept space [15], indicated the feature items of user profile with semantic concepts, calculates learner's interest-level to the topic through establishing the word frequency and utilize the suitable calculation methods, mining the concepts within the user's feedback files and the relationship between concepts, combines user's short-term interests and long-term interests to create user profiles model with semantic concept hierarchy tree and embody the drifting of user profile and improves and

completes the user profiles model consistently on the related feedback mechanism.

In [16], the authors proposed a new approach to User Model Acquisition (UMA) which has two important features. It doesn't assume that users always have a well-defined idea of what they are looking for, and it is ontology-based, i.e., it was dealt with concepts instead of keywords to formulate queries. The first problem is that most approaches assume users to have a well-defined idea of what they are looking for, which is not always the case. They solved this problem by letting fuzzy user models evolve on the basis of a rating induced by user behavior. The second problem concerns the use of keywords, not concepts, to formulate queries. Considering words and not the concepts behind them often leads to a loss in terms of the quantity and quality of information retrieved. They solved this problem by adopting an ontology-based approach.

In [9], the student interactions with the system are monitored and stored in log files. The recorded data are then cleaned and preprocessed (e.g. compute the relative frequency of learner actions, the amount of time spent on a specific action type, the order of navigation etc). Subsequently, these behavioral indicators are analyzed and based on them the system can infer different learning preferences of the student. Finally, the identified learner model is used by the decision-making component to select the most appropriate adaptation actions, in order to provide the student with the educational resources that suit her/his specific needs.

### III. THE PROPOSED SYSTEM ARCHITECTURE

The design idea of adaptive learning system based on creating an ideal learning environment for the learners so that system can provide the adaptive learning support according to the learner's individual differences, and to promote learners to study initiatively, and to achieve the knowledge construction. There are some objectives in the design of adaptive learning system. First, system can provide the adaptive learning content based on the learner's interest and knowledge requirement. Second, system can support the self-directed learning and collaborative Learning. Third, system can help teachers understand the learning process of learners, and adjust the pedagogical activities, and support the learning evaluation. Last, system will support the courses development for staff. Based on these considerations, a new architecture of adaptive learning system is proposed in current paper. It is illustrated in figure 4. According to proposed architecture, our learning system is mainly composed of four processes: Learner's Web Log Analysis, Learner Interest Builder, and Knowledge Requirement Acquiring. These Processes will be explained in the next section.

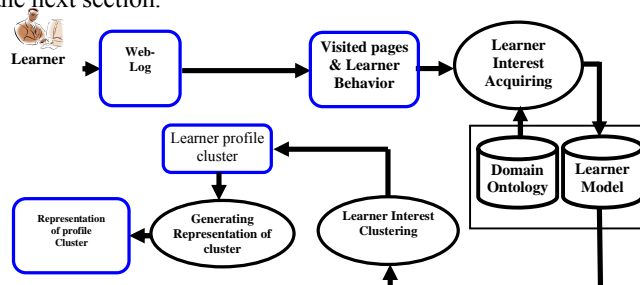


Fig.4. The Architecture of the Proposed

#### A. LEARNER'S INTEREST ACQUIRING

In this system, learner interest model's knowledge expression uses the thought which is based on the space vector model's expression method and the domain ontology [22]. The figure 5 shows certain steps to acquire learner interest.

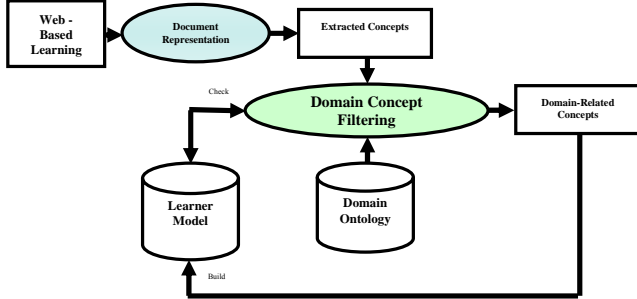


Fig.5: Acquiring E-learners interest Steps

#### • DOCUMENT REPRESENTATION

The Vector Space Model [17,18, 19] is adapted in our proposed system to achieve effective representations of documents. Each document is identified by n-dimensional feature vector where each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight. The weight is a function of the *term frequency*, *collection frequency* and *normalization factors*. Different weighting approaches may be applied by varying this function. Hence, a document  $j$  is represented by the document vector  $d_j$ :

$$d_j = (w_{1j}, w_{2j}, \dots, w_{nj}) \text{ Where, } w_{kj} \text{ is the weight of the}$$

$k^{\text{th}}$  term in the document  $j$ .

The term frequency reflects the importance of term  $k$  within a particular document  $j$ . The weighting factor may be global or local. The global weighting factor takes into account the importance of a term  $k$  within the entire collection of documents, whereas a local weighting factor considers the given document only. Document keywords were extracted by using a term-frequency-inverse-document-frequency (tf-idf) calculation [18, 19], which is a well-established technique in information retrieval. The weight of term  $k$  in document  $j$  is represented as:

$$w_{kj} = tf_{kj} \times (\log_2^n - \log_2^{df_k} + 1)$$

Where:  $tf_{kj}$  is the term  $k$  frequency in document  $j$ ,  $df_k$  is number of documents in which term  $k$  occurs,  $n$  = total number of documents in collection.

Table 1 shows the term frequency in different documents. The main purpose of this step is to extract interested items in the web page, then get term frequency that reflects the importance of term, finally get the weight of terms in the selected page. The output of this step is the weight of terms in selected page that can be used to build learner interest profile.

#### • DOMAIN CONCEPT FILTERING

This process discovers concepts which represent the learner's interests. These concepts and items are compared to the domain ontology to check the relevant items to the learner profile. The most relevant ones update the learner profile. The items relevance is based on ontology-based semantic similarity where browsed items by a learner on the web are compared to the items from a domain ontology and learner profile. The importance is combined with the semantic similarity to obtain a level of relevance. The page items are processed to identify domain-related words to be added to the learner profile. A bag of browsed items is obtained via a simple word indexing of the page visited by the learner. We filter out irrelevant words using the list of items extracted from domain ontology. Once domain-related items are identified, we evaluate their relevance to learner's interests. The selected method was used in [20, 21] to compute semantic similarity function ( $S$ ) based on a domain ontology. The similarity is estimated for each pair of items where one item is taken from a learner profile, while the other one from a set of browsed items. The functions  $S_w$  is the similarity between synonym sets,  $S_u$  is the similarity between features, and  $S_n$  is the similarity between semantic neighborhoods between entity classes  $a$  of ontology  $p$  and  $b$  of ontology  $q$ , and  $w_w$ ,  $w_u$ , and  $w_n$  are the respective weights of the similarity of each specification component.

$$S(a^p, b^q) = w_w \times S_w(a^p, b^q) + w_u \times S_u(a^p, b^q) + w_n \times S_n(a^p, b^q) ;$$

For  $w_w$ ;  $w_u$ ;  $w_n \geq 0$ :

Weights assigned to  $S_w$ ,  $S_u$ , and  $S_n$  depends on the characteristics of the ontologies.

The similarity measures are defined in terms of a matching process [20, 21]:

$$S(a, b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a, b)|A/B| + (1 - \alpha(a, b))|B/A|}$$

where  $A$  and  $B$  are description sets of classes  $a$  and  $b$ , i.e., synonym sets, sets of distinguishing features and a set of classes in semantic neighborhood;  $(A \cap B)$  and  $(A/B)$  represent intersection and difference respectively,  $| \cdot |$  is the cardinality of a set; and  $\alpha$  is a function that defines relative importance of non-common characteristics. A set of browsed items that are similar to items from the user profile is considered as a set of items that can be added to this profile. Table 2 shows a sample of the weighted terms in the documents; that found in table 1.

Table 1: Sample of the Documents with their representation

DOC/items	Computer science	AI	Programming	Software eng.	Networks	LAN	WAN	Computer Arch.	Processors	Parallel processing
Doc1	20	25	20	15	10	0	0	5	0	0
Doc 2	15	15	25	15	5	0	0	0	0	0
Doc 3	0	5	25	10	25	10	15	5	0	0
Doc 4	0	5	25	10	20	5	15	5	0	0
Doc 5	0	0	5	0	10	0	0	5	20	20
Doc 6	0	0	0	5	20	0	0	0	25	30
Doc 7	0	0	5	0	0	0	0	5	20	10
Doc 8	0	0	0	5	0	0	0	10	25	5
Doc 9	15	5	30	30	0	0	0	5	0	0
Doc 10	5	0	25	25	0	0	0	0	0	0
Doc 11	10	0	0	0	10	0	0	25	10	30
Doc 12	10	0	0	0	10	0	0	30	10	25
Doc 13	20	25	5	0	10	0	0	5	0	0
Doc 14	15	15	5	0	5	0	0	0	0	0
Doc 15	0	30	20	15	25	0	5	0	0	0
Doc 16	0	25	25	15	20	0	0	0	0	0

Table 2: Sample of the weighted terms in the documents

	computer Science	AI	programming	Software Engineering	Network	LAN	WAN	Computer Architecture	Processors	Parallel Processing
Doc1	40.00	45.75	28.30	25.17	14.15	0.00	0.00	9.15	0.00	0.00
Doc2	30.00	27.45	35.38	25.17	7.08	0.00	0.00	0.00	0.00	0.00
Doc3	0.00	9.15	35.38	16.78	35.38	40.00	51.23	9.15	0.00	0.00
Doc4	0.00	9.15	35.38	16.78	28.30	20.00	51.23	0.00	0.00	0.00
Doc5	0.00	0.00	7.08	0.00	14.15	0.00	0.00	9.15	48.30	48.30
Doc6	0.00	0.00	0.00	8.39	28.30	0.00	0.00	0.00	60.38	72.45
Doc7	0.00	0.00	7.08	0.00	0.00	0.00	0.00	9.15	48.30	24.15
Doc8	0.00	0.00	0.00	8.39	0.00	0.00	0.00	18.30	60.38	12.08
Doc9	20.00	9.15	42.45	50.34	0.00	0.00	0.00	9.15	0.00	0.00
Doc10	10.00	0.00	35.38	41.95	0.00	0.00	0.00	0.00	0.00	0.00
Doc11	20.00	0.00	0.00	0.00	14.15	0.00	0.00	45.75	24.15	72.45
Doc12	20.00	0.00	0.00	0.00	14.15	0.00	0.00	54.90	24.15	60.38
Doc13	40.00	45.75	7.08	0.00	14.15	0.00	0.00	9.15	0.00	0.00
Doc14	30.00	27.45	7.08	0.00	7.08	0.00	0.00	0.00	0.00	0.00
Doc15	0.00	54.90	28.30	25.17	35.38	0.00	17.08	0.00	0.00	0.00
Doc16	0.00	45.75	35.38	25.17	28.30	0.00	0.00	0.00	0.00	0.00

number of instances was 16, the number of attributes was 10, the number of clusters was 2 and their centers are shown in Table 4

#### IV. THE EXPERIMENTS DESIGN AND RESULTS ANALYSIS

1. *The First Experiment: predicting the e-learners interests using the Simple Fuzzy-KMeans* Where the number of instances was 16, the number of attributes was 10, the number of clusters was 2 and their centers are shown in Table 3.
2. *The Second Experiment: predicting the e-learners interests using the weka.clusterers.XMeans*, Where the
3. *The Third Experiment: predicting the e-learners interests using the fuzzy K-Means and attributes selection using weka.attributeSelection.BestFirst.*

Table 3: Results of the first experiment

The learners group	Computer science	AI	Programming	Software eng.	Networks	LAN	WAN	Computer Arch.	Processors	Parallel processing
Cluster 0 number of cases belong to it are 10 ( 63%)	9	15	20.5	13.5	12	1.5	3.5	2.5	0	0
Cluster 1 number of cases belong to it are 6 ( 38%)	3.3	0	1.66	1.66	8.33	0	0	12.5	18.33	20

Table 3: Results of the Second experiment

The learners group	Computer science	AI	Programming	Software eng.	Networks	LAN	WAN	Computer Arch.	Processors	Parallel processing
Cluster 0 number of cases belong to it are 14 ( 88%)	7.8	10	11.7	8.9	8.9	0	0.35	6.4	7.8	8.5
Cluster 1 number of cases belong to it are 2 ( 13%)	0	5	25	10	22.5	7.5	15	5	0	0

Table 4: Results of the Third experiment (*The number of selected attributes are: 5*)

The learners group	AI	Programming	Computer Arch.	Processors	Parallel processing
Cluster 0 number of cases belong to it are 10 cases 63%	15	20.5	2.5	0	0
Cluster 1 number of cases belong to it are 6 cases 38%	0	1.666	12.5	18.333	20

## V. SUGGESTIONS AND RECOMMENDATIONS

The Semantic Web can be used to organize information in concept structures, while web services allow the encapsulation of heterogeneous knowledge and modularization of the architecture. The key idea of the Semantic Web is to have data defined and linked in such a way that its meaning is explicitly interpretable by software processes rather than just being implicitly interpretable by humans.

The user profiles can maintain sophisticated representations of personal interest profiles. These representations can be utilized for effective information retrieval. Fuzzy clustering allows an entity to belong to more than one cluster with different degrees of accuracy, while hard clustering assigns each entity exactly to one of the clusters. Thus, fuzzy clustering is suitable in constructing the learner profiles representations such as (learner ontology). Such representation of user profiles is useful because some information is not forced to fully belong to any one of the user profiles. Fuzzy clustering methods may allow some information to belong to several user profiles simultaneously with different degrees of accuracy.

Our proposed approach depends on using semantic web to extract the learner model. Fuzzy technique is used to cluster the extracted data of learner model to classifying the learners for their interests. This classifying the interests of learners enable learning systems to handle learner as groups to recommend them what they must teach corresponding to their interests in the learner profile clusters.

Cluster of learner profiles must be used in e-learning system, so we recommend to represent the cluster of profiles on learner ontology to express sophisticated user profiles in order that a user may own several interests. A user profile may

correspond to the several ontologies. Ontology-based user profiles typically maintain sophisticated representations of personal interest profiles. These representations can be utilized for effective information retrieval in the e-learning systems.

## VI. CONCLUSION

In this paper, we presented a method for constructing learner model that represents the user's interests by analyzing the web-log to extract the interested terms in visited pages by learner. Then the fuzzy clustering is used to extract clusters of output learner profiles. The goal of incorporating the semantic web is to build the semantically enhanced user models. Fuzzy technique is used to cluster the extracted data of learner model to classifying the learners for their interests. We recommend to use Ontology-based user profiles to maintain sophisticated representations of personal interest profiles.

## ACKNOWLEDGMENT

The authors wish to thank Taif University for the support of this research and the anonymous reviewers for their valuable comments.

## REFERENCES

- [1] D. Vldan. (2006), SEMANTIC WEB AND EDUCATION, ISBN: 0-387-35417-4, Springer Science+Business Media, LLC.
- [2] F. Christoph, (2005). User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis At Graz University of Technology.
- [3] R. Marek and K. Sayed. (2009), Updating User Profile using Ontology-based Semantic Similarity, FUZZ\_IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.
- [4] A. YACINE, B. RACHID, B. JAWAD. (2004), Learning Objects Based Framework for Self-Adaptive Learning, Education and Information

- Technologies 8:4, 345–368, 2003, 2004 Kluwer Academic Publishers. Manufactured in the Netherlands.
- [5] S. Uri, E. Masha. (2006), e-Learning for depth in the Semantic Web, British Journal of Educational Technology, Vol 37 No 3 2006 425–444, Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.
  - [6] K. Breitman, M.Casanova and W. Truszkowski. (2007), Semantic Web: Concepts, Technologies and Applications, e-ISBN: 978-1-84628-710-7, Springer-Verlag London Limited.
  - [7] D. Dicheva. (2008), Ontologies and Semantic Web for E-Learning, in book "Handbook on Information Technologies for Education and Training", 978-3-540-74155-8 (Online), Springer Berlin Heidelberg.
  - [8] G. Fayed, D. Sameh, H. Ahmad, M. Jihad., A. Samir, and S. Hosam. (2006), E-Learning Model Based On Semantic Web Technology, International Journal of Computing & Information Sciences Vol. 4, No. 2, August 2006, On-Line. Pages 63 – 71.
  - [9] P. Elvira, T. Philippe, B. Costin.(2007), Adaptive Educational Hypermedia Systems: A Focus on Learning Styles, EUROCON 2007 The International Conference on "Computer as a Tool" Warsaw, September 9-12, 1-4244-0813-X/07, IEEE.
  - [10] LTSC Learner Model Working Group of the IEEE (2000), IEEE p1484.2/d7, 2000-11-28 Draft Standard for Learning Technology - Public and Private Information (PAPI) for Learners (PAPI Learner), Technical report Available Online: [http://www.edutool.com/papi/papi\\_learner\\_07\\_main.pdf](http://www.edutool.com/papi/papi_learner_07_main.pdf)
  - [11] Smythe, C., F. Tansey, R. Robson (2001), IMS Learner Information Package Information Model Specification, Technical report Available Online: <http://www.imsglobal.org/profiles/lipinfo01.html>.
  - [12] 12. [12] F. Christoph. (2005), User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis at Graz University of Technology, Copyright 2005 by Christoph Fröschl.
  - [13] D. Trong, N Mohammed, L. Delong , and J. Geun. (2009), A Collaborative Ontology-Based User Profiles System, N.T. Nguyen, R. Kowalczyk, and S.-M. Chen (Eds.): ICCCI 2009, LNAI 5796, pp. 540–552, Springer-Verlag Berlin Heidelberg.
  - [14] B. Qiu, W. Zhao. (2009), Student Model in Adaptive Learning System based on Semantic Web, In First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466.
  - [15] W. Cuncun, H. Chongben, T. Hengsong.(2009), A Personalized Model for Ontology-driven User Profiles Mining, 2009 International Symposium on Intelligent Ubiquitous Computing and Education, 978-0-7695-3619-4/09, IEEE.
  - [16] C'elia and T. Andrea G.B. (2006), An Evolutionary Approach to Ontology-Based User Model Acquisition, V. Di Ges'u, F. Masulli, and A. Petrosino (Eds.): WILF 2003, LNAI 2955, pp. 25–32, c\_Springer-Verlag Berlin Heidelberg.
  - [17] B. Qiu, W. Zhao. (2009), Student Model in Adaptive Learning System based on Semantic Web, 2009 First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466
  - [18] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng, and W. Daming. (2007), Ontology Based User Profiling in Personalized Information Service Agent, Seventh International Conference on Computer and Information Technology, 0-7695-2983-6/07, IEEE.
  - [19] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng. (2007), A Personalized Semantic Search Method for Intelligent e-Learning, 2007 International Conference on Intelligent Pervasive Computing, 0-7695-3006-0/07, IEEE , DOI 10.1109/IPC.2007.48.
  - [20] R. Marek and K. Sayed. (2009), Updating User Profile using Ontology-based Semantic Similarity, FUZZ\_ IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.
  - [21] M. Andrea, E. Max J. (2003), Determining Semantic Similarity among Entity Classes from Different Ontologies, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 15, NO. 2, MARCH/APRIL, 1041-4347/03, IEEE.
  - [22] H. Hany, F. Kh. (2010), "Semantic web based Approach to learn and update Learner Profile in Adaptive E-Learning ", Al-Azhar Engineering Eleventh International Conference, December 23-26, pp., 2010.
  - [23] M. Oded, R. Lior. (2010) , Data Mining and Knowledge Discovery Handbook Second Edition, In chapter " A survey of Clustering Algorithms " ISBN 978-0-387-09822-7, Springer Science+Business Media.

#### AUTHORS PROFILE



**Khaled M. Fouad** received his Master degree of AI and expert systems. He is currently a PhD candidate in the faculty of engineering AlAzhar University in Egypt. He working now as lecturer in Taif university in Saudi Arabia and is assistant researcher in Central Laboratory of Agriculture Expert Systems (CLAES) in Egypt. His current research interests focus on Semantic Web and Expert Systems.



**Mofreh A. Hogo** is a assistant professor at Benha University, Egypt. He is a assistant professor of Computer Science and Engineering. Dr. Hogo holds a PhD in Informatics Technology from Czech Technical University in Prague, Computer Science and Engineering Dept. 2004. He is the author of over 40 papers that have been published in refereed international Journals (Information Sciences, Elsilver, UBICC, IJICIS, IJCSIS, IJPRAI, ESWA, IJEL, Web Intelligence and Agent Systems, Intelligent Systems, international journal of NNW, IJAIT Journal of Artificial Intelligence Tools, IJCI).



**Shehab Gamalel-Din** is Full Professor in King Abdul-Aziz University, Faculty of Science, Computers Dept. Jeddah, KSA and Al-Azhar University, Department of Computers and Systems Engineering Cairo, Egypt. He has Professional Memberships in Institute of Electrical and Electronics Engineers (IEEE), USA, Association of Computing Machinery (ACM), USA, and Egyptian Society for Engineers, Egypt.



**Nagdy Mahmoud Nagdy** is Professor of engineering applications and computer systems, Department of Systems Engineering and Computer Engineering - Faculty of Engineering Al-Azhar University. He received his PH.D in 1986. He has Supervision of some master's and doctoral degrees in the Department of Systems Engineering and Computer and Electrical Engineering.

# An Innovated Server-Based Desktop Sharing Platform

Ayad Hussain. Abdulqader<sup>1</sup>, Salah Noori Saleh<sup>1</sup>, Omar Amer Abouabdalla<sup>1</sup>, Sureswaran Ramadas<sup>1</sup>, Hasan S. M. Al-Khaffaf<sup>2</sup>

<sup>1</sup>National Advance IPv6 Centre of Excellence (NAv6)

USM 11800, Penang, Malaysia

{ayad | salah | omar | sures} @nav6.org

<sup>2</sup>School of Computer Sciences

USM 11800, Penang, Malaysia

hasan@m.ieice.org

**Abstract-** In this paper, a server-based Desktop Sharing Platform (DSP) is proposed. The proposed platform is designed to work with any direct-connection Remote Desktop System (RDS) without modifying or adding any extra models to those systems' limitations in terms of bandwidth consumption, collaboration session initiation, and connectivity issues will be overcome by adopting the proposed platform. The proposed platform is easily adapted to work with any direct-connection RDS system. Incorporating the proposed platform will improve the performance and efficiency of existing RDS systems. As a result, better utilization of computer system resources in terms of bandwidth and processing power is achieved by minimizing the data transfer and processing power from n users to only one user.

**Keywords-** Computer Supported Cooperative Work; Remote Display System; Thin-Client Computing

## I. INTRODUCTION

A great part of a people's work time is consumed by collaborative tasks. It can be a challenge to collaborate effectively in the absence of the communication tools to share information and coordinate project details [1]. Therefore, many real-time interactive systems use communication technology for supporting collaborative work. These systems use a computer display as a medium to enable such group of users to be updated on the current status of certain work. One of these systems is the Remote Desktop System (RDS), which is the main focus of this paper. Relying on RDS systems will enable users to access their PCs from anywhere in the world and a group of users can share remote PCs' resources. In RDS a Remote Display Protocols (RDP) is used to convey data among different users by transferring graphics display information from the remote computer (server) to the connected users (clients) and transporting input events from those users to the remote computer.

There are numerous studies investigating the design of new protocols for RDSs. Most of these studies concerned with enhancing the performance of these systems by inventing new methods of conveying output display and minimizing the quantity of the sent output display data. However, these studies lack the concern of leveraging these systems to be used effectively in a collaboration environment.

## II. BACKGROUND

In the past decades, the fast development of computer hardware and software technologies, especially the explosive expansion of Internet, has boosted research on Computer Supported Cooperative Work (CSCW) and groupware techniques. Numerous groupware systems have been developed as commercial products and research prototypes. Groupware systems can be classified, according to the time and date dimensions, into four categories as follows [2, 3]:

- **Face to Face Interaction:** Participants meet face-to-face in one place at the same time, as in a traditional meeting.

- **Asynchronous Interactions:** Participants are in the same place but they work at different time. For example, people work in shifts. One shift leaves information for the next shift.

- **Synchronous Distributed Interaction:** Participants are in different places but they communicate at the same time, for example video conferencing.

- **Asynchronous Distributed Interaction:** Participants are in different places and work at different times for example electronic mail and bulletin board systems.

Two architectures are usually considered in designing groupware systems, the replicated architecture and centralized architecture [4]. In replicated architecture, each collaborating site has an instance of the shared application running at the local site. By adopting this architecture the network bandwidth is improved, where only input events need to be broadcasted to all sites. But the consistency of the data is more difficult to guarantee than central architecture because multiple copies of data exist. This considered as the main disadvantage of this architecture.

In centralized groupware architecture there is only one instance of the shared application runs on a single host. Other collaborating sites have only client end systems with limited functions. All user inputs to the shared application are forwarded to this single instance while the outputs are broadcasted back to different client ends. RDPs are usually used to convey the data exchanging between the central host and all connected clients. Broadcasting the outputs to the all collaborating sites will lead to inefficient use of network bandwidth as well as a heavy processing incurred at central host.



### III. REMOTE DESKTOP SYSTEMS

An RDS consists of a server and a client that communicate over a network using RDP. The protocol allows graphical displays to be virtualized and served across a network to a client device, while application logic is executed on the server. Using the remote display protocol, the client transmits user input to the server, and the server returns screen updates of the user interface of the applications to the client.

RDS can provide many potential benefits, including [5]:

- Remote Desktop enables a user to work remotely with any remote PC.
- Full graphical access to use computer-controlled scientific instrumentation remotely.
- Supporting a desktop sharing among multiple users to support online collaboration.
- Giving the ability of fixing a software problem remotely.
- Moving application state off the desktop back into the machine room such that clients only need to serve as simple remote display devices, thereby reducing IT management complexity via the centralization benefits that come from a thin-client computing approach.

There are many remote display protocols defined by different vendors or organizations, such as Remote Frame Buffer (RFB) for Virtual Network Computing (VNC) [6], Remote Desktop protocol for Windows Terminal Service, ITU-T T.128 for NetMeeting and SunForum [7], and a Red Hat spice protocol for spice project [8]. These protocols work in two phases: the setup phase and a session phase. At setup phase a negotiation between the server and the client on setting the initial values of the session phase parameters. As in RFB protocol, the parameters include pixel format and/or the security and compression types. In the session phase the graphic data are being sent by the server to the client, and the mouse and keyboard operations are being sent from the client to server [6, 9].

RDSs suffer from a series of issues when it is used in collaboration environment. RDS Server in these systems distributes data individually and directly to the connecting participants (or RDS clients), as shown in fig. 1.

Due to the centralized distribution architecture of RDS systems, RDS server should broadcast the output display information to all connected RDS clients when any changes happened to the output display. This broadcasting of display information leads to high demand use of network bandwidth. This becomes more severe with the increasing number of RDS clients. The processing power is also increased with the increasing number of RDS clients as each RDS client needs a special capturing and encoding process. This means that both resources in RDS server (network bandwidth use and the processing power) are directly proportional to the number of RDS clients. This also will affect the overall performance in RDS server that reflected inefficiently on the quality of the desktop sharing at RDS clients' side.

In RDS systems, RDS clients may not be able to always communicate directly and exchange data with the RDS server due to network limitation (e.g., disjoint networks or limited network capacity), and connectivity problems (e.g., uncooperative service providers or security issues).

Furthermore, with these systems users are lacking the connectivity platform to start and invite each other to the session. Finally there is no unified desktop sharing platform that has the ability to work effectively with any remote desktop systems.

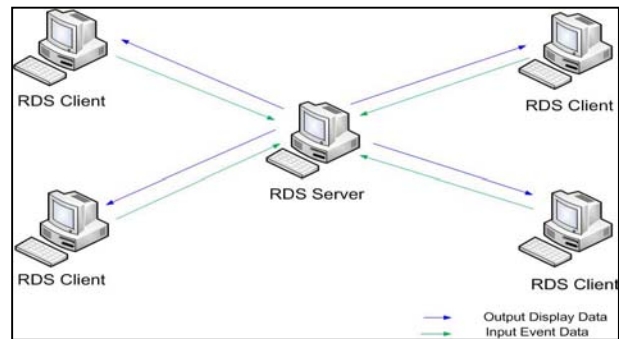


Figure 1. Direct-Connection Desktop Sharing Environment

There are two categories of messages exchanged between the RDS server and the RDS client, setup messages and session messages. The setup messages are exchanged between the RDS server and the RDS client during a setup phase to determine the session phase parameters. During the session phase, graphic data messages and input events messages are exchanged between RDS server and RDS client. Fig. 2 shows a message exchange process during the setup and session phases.

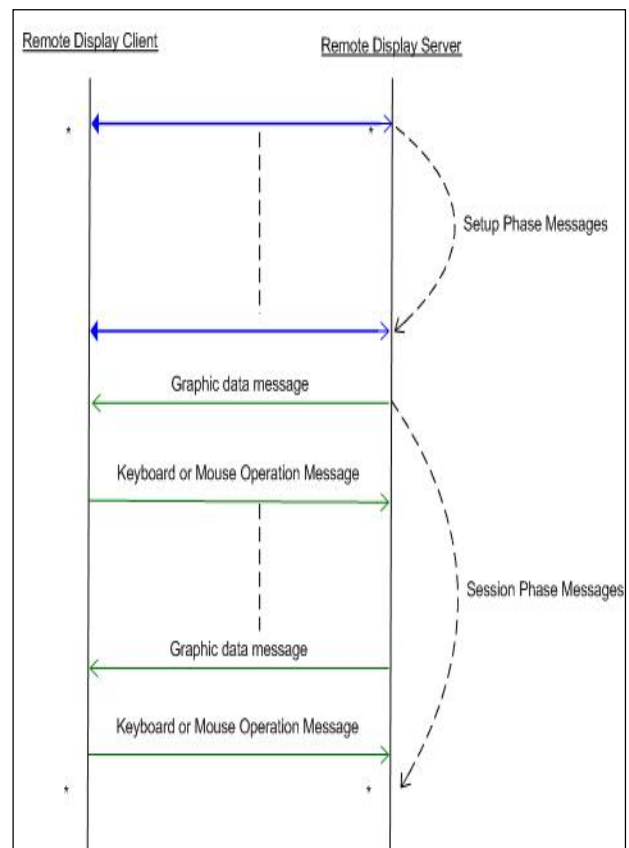


Figure 2: The two phases of the Remote Display Protocol.

#### IV. THE PROPOSED DESKTOP SHARING PLATFORM (DSP)

The proposed server-based Desktop Sharing Platform (DSP) will eliminate all of the above limitations and leverage existing RDS systems to work effectively in a collaboration environment. The DSP is designed to be adapted (configured) easily to work with any RDS system that uses TCP as a transport protocol without adding or modifying any module to both the RDS server and the RDS client sides. In other words, configuring the proposed DSP platform with the configuration data of the RDS system  $x$  will produce the DSP platform for RDS system  $x$ , as follows:

$$DSP(x) = \text{Un-configured DSP} + \text{ConfigurationDataOf}(x)$$

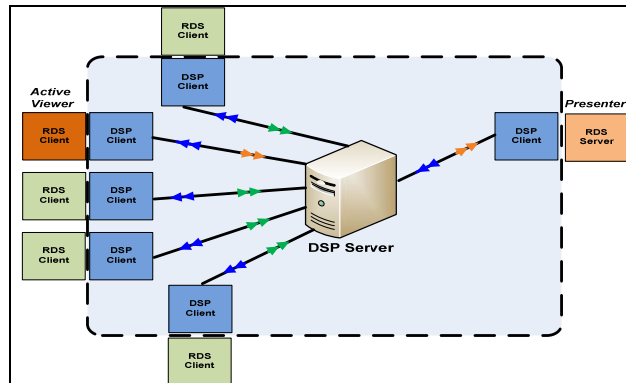


Figure 3. The Proposed Server-Based Desktop Sharing Platform (DSP)

The proposed DSP platform is a client-server system, as shown in fig. 3, where the DSP server is responsible of managing a collaboration session among connected DSP clients beside analyzing, buffering and distributing RDS server/clients' messages.

The collaboration session participants in our proposed platform are classified into two categories, presenter and viewers. There is only one presenter and one or more viewers during a certain collaboration session. Each viewer consists of an RDS client and its corresponding DSP client. The RDS client is a client of any direct-connection RDS, e.g. VNC client.

The DSP client is a thin-client that is located with RDS server/client on the same machine. The DSP clients are responsible of handling all collaboration messages with a DSP server along with controlling and managing the exchanged data between the RDS server/client and DSP server during a collaboration session. There are two types of DSP clients in our proposed DSP platform; the first type is located at the RDS server, see *presenter* in fig. 4, whereas the second type is located at the RDS client, see *viewers* in fig. 4.

The DSP client for RDS server acts as a client to the RDS server. In this case, all the messages from RDS server are sent to its corresponding DSP client. The DSP client, in turn, will pass RDS server's messages to the DSP server. The DSP client for RDS client acts as a server to the RDS client, so that the entire RDS client's messages are sent to its corresponding DSP client then to the DSP server.

The exchanging messages between the RDS client and the RDS server will go through the DSP server then through their corresponding DSP client. The presenter participant module consists of an RDS server and its corresponding DSP client. The RDS server is a server of any direct-connection RDS, e.g. VNC server. Typically one participant, which is called the active viewer, can be granted with a floor control feature at a certain time compared to the viewing status of the other viewer participants. Within a desired session, a participant who invites other users to establish a session is called a chairman. A chairman has the authority to assign floor control feature to the specified participant (choose active viewer), to assign a presenter feature to the specified participant, and to end the session.

In DSP platform the data distribution behavior among RDS server and its clients is changed from multiple direct connections to the single connection through a DSP server. In this case, the RDS server will send the output display data, through its corresponding DSP client, to the DSP server only. The DSP server, in turn, will broadcast these data to all connected RDS clients through their corresponding DSP clients. The DSP server also collects the input events from presenter clients and passed it to the RDS server.

There are two main functions of the DSP server, establishing and managing a collaboration session, and distributing the RDS messages. Therefore, there exist two types of data flow messages, the control data flow messages and RDS data flow messages, see Fig. 4. The control messages are handled by DSP server and their clients to initiate and manage a collaboration session. The control messages are divided into two groups: client-to-server messages and server-to-client messages. Another function of the DSP server is to store the exchanging messages between the RDS server and its active client to a specified storage. The stored messages can be used by the new comer to view the previous collaboration activities since the beginning of the session as a VCR-like control.

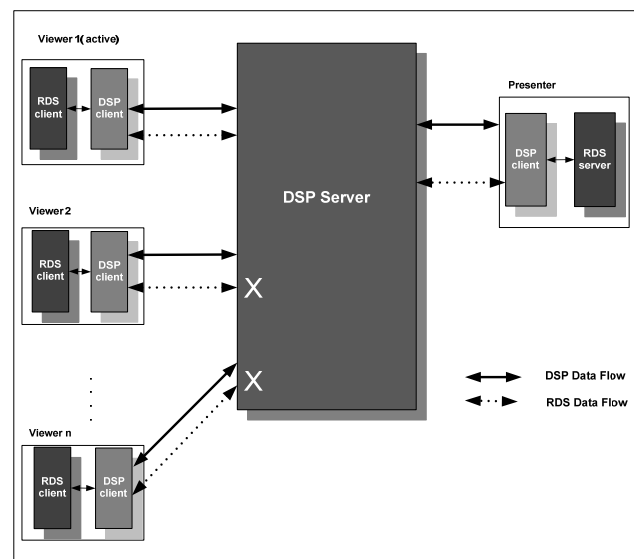


Figure 4. The Data Flow of the Proposed DSP Platform

The *Client-to-Server* control messages are sent from *DSP* client to *DSP* server. Those messages are used to establish connection, invite other participants to join the session, leave from the session, and so on. The following is a list of all these messages with their functions:

**1. Connect [username][password]**

Client initiates a connection (login) to the server, the server will perform verification on username and password then it will send *connAck* message to the client.

**2. Invite [Protocol][userList]**

Client invites users to start a session using specified protocol. A client who sent this message will become the chairman. In turn, the server will use its *invite* message to pass this invitation to all users to get their acceptance.

**3. Close**

Chairman requests the server to close the current session. At this moment the server will inform the clients that the session will go down according to chairman request.

**4. Disconnect**

Client requests to close a connection (logout). If the client is already in the session then the server will close all the client sockets and use *LeftSession* message to inform other participants regarding this disconnection.

**5. Accept [chairman][Caps]**

When a server sent an invite message to the client as a result of chairman invitation, a client will accept or reject this invitation by sending *Accept* message with the chairman's name. The client can choose from a number of capabilities namely: viewer only, presenter only, both, or no (rejects this invitation).

**6. Assign [pres][active]**

This message is sent only by the chairman to inform the server that *pres* participant will become a presenter and *active* will become an active viewer.

**7. PresenterAck**

When a *DSP* client receives an *Assign [presenter]* message it will prepare its *RDS* server to be ready for accepting *RDS* clients' connections.

**8. ActiveAck**

If the client receives *Assign [active]* message, it will reply by sending this message.

**9. UserlistReq**

A chairman client sends this message to get a list of currently connected user before sending the *invite* message.

The *Server-to-Client* messages are sent from *DSP* server to *DSP* client. Those messages along with *Client-to-Server* control messages are used to manage the connection and session initiation for a group of participant. The *Server-to-Client* messages and their functions are listed as follows:

**1. Ack**

A response to the *connect* message to inform the client that the connection is done properly or accordingly.

**2. NotAck[reason]**

A response to the *connect* message to inform the client that the connection is not established.

**3. Invite [Protocol][chairman]**

This message is sent to all users that are invited to establish a session to inform them about this invitation including the usage protocol.

**4. ClientsCapabilities [userListCapabilities]**

As a result of *Accept* message the server collects the invitation acceptances from the clients and passes to the chairman with their capabilities.

**5. LeftSession [username]**

This message is sent when the server receives a *disconnect* message from a specified participant. If a participant is a chairman then the whole session will go down by sending *closeSession* message.

**6. SessionCreated**

This message is sent to all participants including the chairman to inform them that the session is created and the server is ready to manage and distribute messages among them. It holds the number of connection channel along with the addresses of the participant agents. The *DSP*'s addresses are supplied to each participant to enable its *DSP* client to initiate a connection to the *DSP* server.

**7. CloseSession**

This message is sent to all participants informing them that the current session will go down.

**8. Assign [presenter| active]**

This message is sent to a specified participant informing him that he will be a presenter or an active viewer.

**9. UserListResp**

Server responds to the *UserListReq* by sending the list of the currently connected user.

## V. COLLABORATION SESSION INITIATION

To describe a connection establishment and initiation of collaboration session processes consider for example four presumed clients (*A*, *B*, *C*, and *D*). Fig. 5 shows the connection process where each client sends its user ID and password with the *connect* message. When the server verifies the user ID and password, it sends *Ack* message to the client for acknowledgment.

The connected clients can now invite each other to establish a collaboration session. Fig. 6 shows the scenario when client *B* wants to establish a session with clients *A*, *C*, and *D*. Client *B* sends an *UsrLstReq* message to the server. The server responds with *UsrLstRes* informing client *B* that clients *A*, *C*, and *D* are currently connected to the server. Client *B* uses *Invite* message to invite clients *A*, *C*, and *D* to establish session using Remote Frame Buffer protocol (RFB). The server in turn, passes client *B* invitations to the invited clients using *Invite* message. Now each invited client sends the *Accept* message to inform the server its capabilities.

There are four capabilities: *presenter only*, *viewer only*, *both*, and *no*. In the presumed scenario shown in Fig. 6, clients **A**, **C**, and **D** have the capabilities *both*, *viewer*, and *no* respectively, *both* means that the client **A** can be *presenter* and/or *viewer*, and *no* means that the client **D** does not support RFB protocol.

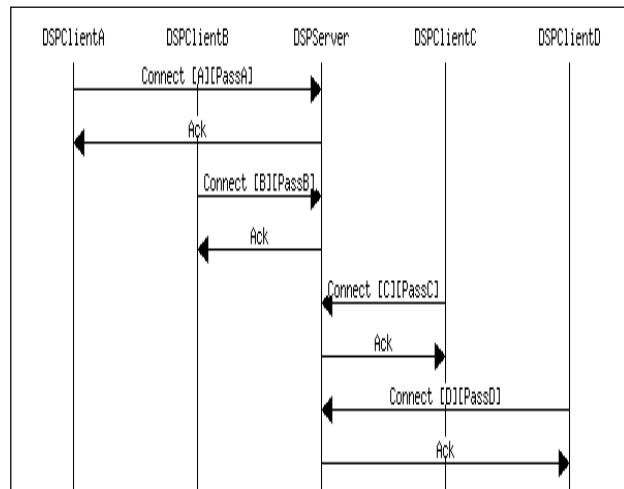


Figure 5. The Connection scenario between DSP server and four DSP clients (DSP client A, DSP client B, DSP client C, and DSP client D)

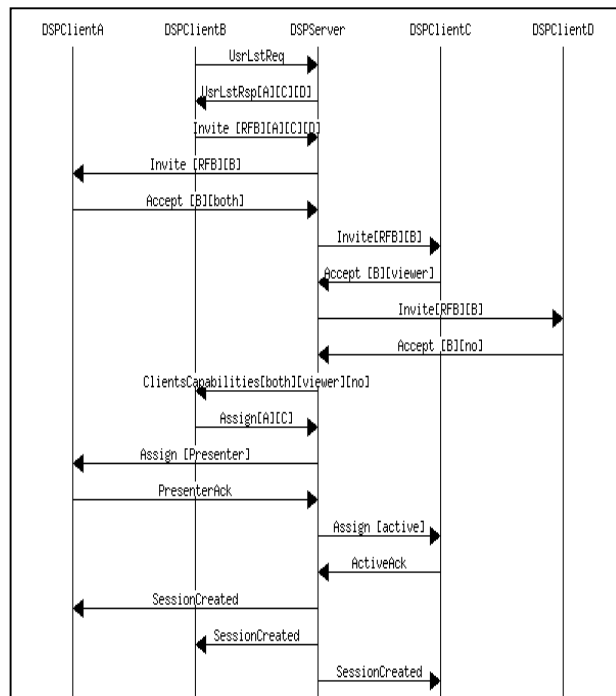


Figure 6. A scenario of establishing a collaborative session among DSP client A, DSP client C, and DSP client D through the DSP server

After collecting information of clients' capabilities, the server passes them to the client **B** (Chairman) by sending the *Acceptance* message. Now the chairman should assign *presenter* and *active* features to the desired clients so that

client **B** sends an *Assign[A][C]* message to tell the server that the presenter is client **A** and the active viewer is client **C**. In turn, the server sends *Assign[Presenter]* message to client **A** and *Assign[Active]* to client **C**.

A *PresenterAck* and *ActiveAck* are sent from the presenter and the active viewer respectively informing the server that they are ready. In turn, the DSP server broadcasts *SessionCreated* message informing the remaining clients that the presenter is ready.

## VI. THE DISTRIBUTION BEHAVIOR OF RDS MESSAGES THROUGH DSP PLATFORM

In our design of proposed DSP platform, the distribution behavior of RDS messages is divided into two phases the setup phase and the session phase. During the setup phase the DSP server passes the setup messages between the RDS server and the first RDS client (RDS client 0). In turn, the DSP server extracts the setup parameters that are used later by DSP server, in serving the remaining RDS clients requests instead of the RDS server.

During the session phase, the DSP server receives the output display messages from the RDS server through its corresponding DSP client and broadcasts them to all the connected RDS clients through their corresponding DSP clients. Also the DSP server receives input events' messages from the connected RDS clients and passes only those messages coming from the active viewer participant. Fig. 7 illustrates the distribution behavior of RDS messages among joined participants.

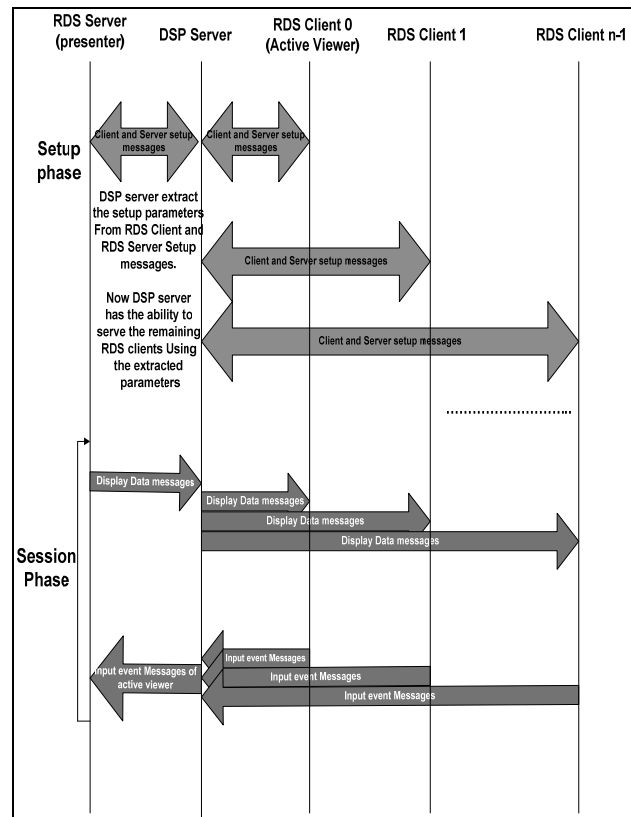


Figure 7. The distribution behavior of RDS messages through DSP Server

## VII. THE MESSAGE DETECTION FINITE STATE MACHINE

The proposed DSP server is designed to be used with any RDS systems that use a TCP as a transport protocol. The distribution behavior of the proposed DSP platform mainly depends on the extracted parameters from setup messages. Extracting those parameters requires first to extract the setup messages themselves. Because of the byte-stream nature of the TCP, detecting RDS message boundaries is considered as an impossible job unless the DSP server is developed for a specified RDS protocol but this will contradict the main objective of the research in reducing the developing efforts to adapt this platform working with any RDS systems. Therefore, a detector of message boundaries can be utilized. The key point of knowing the message boundaries is by detecting the first byte of the message and its length. Some protocols have a specified field in their messages' structure that determines the bytes length of the message and some others protocols do not have such field. In case the protocols have a specified field length, detecting the message boundaries is easy done by reading the value of the length field to know the end boundary of the current message and the start boundary of the next message. However, when the protocols do not have such field length, detecting the message boundaries is imperative in order to execute a specified action or act according to the specified condition. To satisfy the main objective of the proposed platform, we propose a new Finite State Machine, called Message Detection Finite State Machine (MD-FSM). The proposed MD-FSM has the ability to detect messages of any protocol just by feeding the MD-FSM with the profile of the desired protocol.

## VIII. THE DESIGN ARCHITECTURE OF THE PROPOSED DSP PLATFORM

Fig. 8 shows the basic components of the proposed DSP platform of its client and server. The following subsections explain the functions of these components and relationship among them.

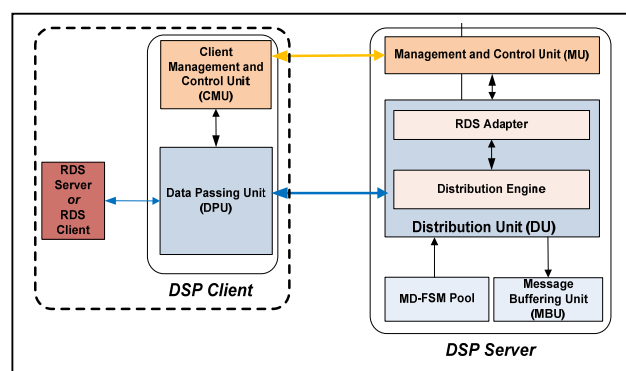


Figure 8 The basic architecture of the proposed DSP platform

### A. The DSP server components

The DSP Server consists of four components; The Management Unit (MU), The Distribution Unit (DU), The Message Buffering Unit (MBU), and the MD-FSM Pool.

### 1. The Management Unit (MU)

The management unit is responsible for managing the collaboration session. So it handles all the messages between the DSP server and its clients. It collects information about joined participants, including, the number of participants and their capabilities, and which RDS system will be used. This information should be provided to the DU in order to prepare its components to do the data distribution. Then the MU collects from the DU the required information to enable each DSP client to establish a direct connection to its corresponding agent in the DU.

### 2. The Distribution Unit (DU)

The DU consists of two modules; an RDS adapter and a Distribution Engine (DE):

#### i. The RDS Adapter

One of the main objectives of this research is to design a collaborative platform that is easily adaptive to work with any RDS systems. As mention above, all RDS systems work in two phases the setup phase and the session phase. In the setup phase, a negotiation occurs between the RDS server and the RDS client to setup the session phase parameters. In our design of DSP platform, the setting up of parameters is done only between the RDS server and the first RDS client. The setting up of parameters for the other clients are accomplished (using the same setup parameters for the first client) by the RDS adapter instead of the RDS server. Setting the session phase parameters might be different from one RDS system to another, for this reason the RDS adapter should be developed for each RDS system to overcome this problem. Thus, the RDS adapter should intercept the exchanging messages between the RDS server and the RDS client to extract the setup parameters. This requires from RDS adapter to detect the exchanging messages first and then extract the needed parameters. To enable the RDS adapter to detect messages, the developer should develop an RDS server simulator within the RDS adapter. However, this will make the RDS adapter more complex to develop and it will contradict the objective of this research. Thus, to overcome this issue and to make RDS adapter simple and easy to develop, we reassigned the message detecting job from the RDS adapter to the distribution engine unit by using the proposed MD-FSM.

As a result, the RDS adapter is now easy to develop and its function can be summarized by the following:

- Extracting the setup parameters from the exchanging setup phase messages between RDS server and the first joined participant (RDS client). The messages themselves are detected by the distribution engine and then sent to the RDS adapter to do the extracting. Fig. 9 shows the flowchart for the function behavior of the RDS adapter.
- Instead of the RDS server and depending on the extracted setup parameters, the RDS adapter becomes responsible for handling the RDS setup phase messages of the new comer participants.



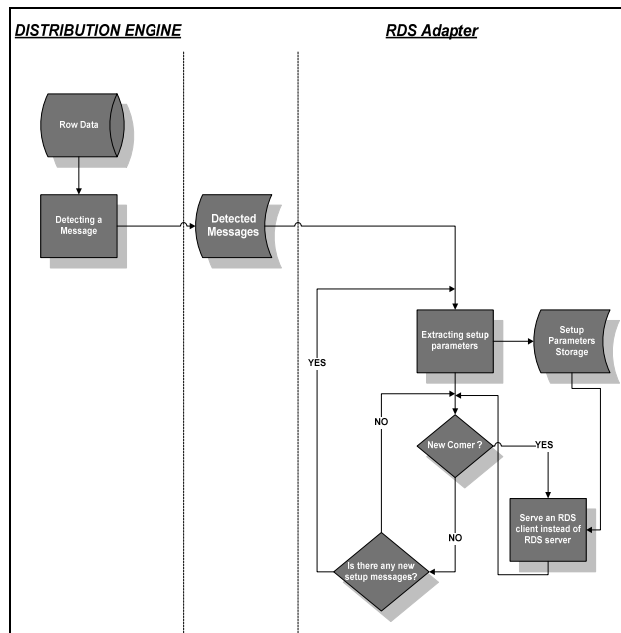


Figure 9: Detecting messages and extracting setup parameters

## ii. The Distribution Engine

The distribution engine consists of one or more of the distribution objects. The number of distribution objects is determined by the number of connection channels between the RDS server and each of the RDS clients. In this way, for each channel there is a Distribution Object (DO) that responsible for handling the distribution of data for this channel. Some RDS systems use one channel (one TCP connection) during a session, whereas others might use more than one channel. The distribution object distributes the data coming from the presenter (RDS Server) to the viewers (RDS Clients) and transfers the data coming from the active viewer to the presenter. Fig. 10 shows a diagram for the distribution engine with  $m$  channels and  $n$  viewers. The diagram shows how  $DO_0, DO_1, \dots, DO_{m-1}$  distribute the data provided from  $channel_0, channel_1, \dots, channel_{m-1}$  to  $viewer_0, viewer_1, \dots, viewer_{n-1}$  respectively

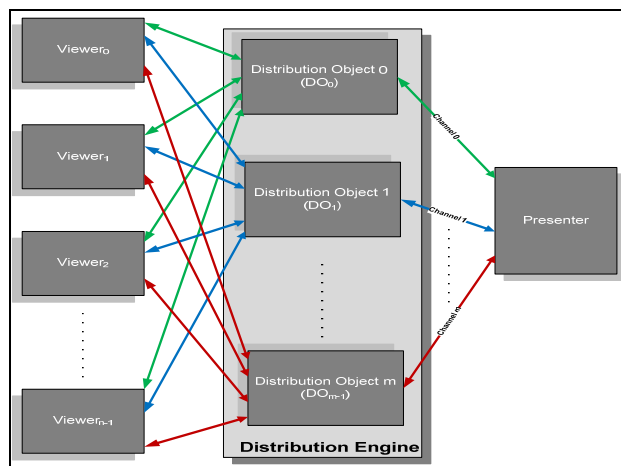


Figure 10: The Distribution Engine

## 3. The Message Buffering Unit(MBU)

The MBU contains the Message Buffer (MB) that holds all generated messages from MD-FSMs. A linked queue data structure is used in designing the MB. The MB is a common buffer between the RDS adapter and all MD-FSMs. The MD-FSMs generate messages and en-queue them to the MB. The RDS adapter de-queues all messages from MB. It extracts the session phase parameters from a setup messages and eliminate the other messages.

## 4. The MD-FSM Pool

The MD-FSM pool holds the MD-FSMs' internal representation of RDS systems that the proposed DSP platform configured to work with them before. Depending on the contents of this pool, the MU decides to agree to establish a collaboration session when the MD-FSM of the specified RDS system already stored in the pool, otherwise it rejects to establish this session.

## B. The DSP Client Components

There are two types of DSP clients. The first one is for the presenter participant that is located with the RDS server at the same machine, and it is called DSP/P, and the second one is for viewer participant that is located with RDS client at the same machine, and is called the DSP/V. Both DSP clients carry out two jobs,

- Handling the connection and collaboration session messages with the MU of DSP server.
- Passing the data coming from RDS client and RDS server to the distribution engine of the DSP server and vice versa.

According to their jobs, the DSP client components are divided in two main units: The Client Management Unit (CMU) and the Data Passing Unit (DPU). The internal structure of the DSP client is shown in Fig. 8.

The CMU handles all the connection and collaboration messages with the DSP server. It gets the TCP socket addresses of all corresponding agents in DSP server's distribution engine from the DSP server. These addresses will be used by the DPU to initiate a connection to these agents and pass data received from the RDS client/ server and vice versa.

The DPU holds one or more Data Passing Object (DPO). The number of DPOs is the same as the number of RDS connection channels. There are two types of DPOs, the DPO for Presenter (DPO/P) and DPO for viewer (DPO/V). The DPO/P acts as double client to both RDS server and its corresponding agent in the DSP Server. The DPO/V acts as a server to the RDS client and as a client to its corresponding agent in the DSP server. Both DPO/P and DPO/V pass the data coming from DSP server to the RDS server/client and vice versa.

## IX. RESULTS AND DISCUSSION

According to our design of the DSP platform, the DSP platform is a client-server system. Both the client and server consist of independent units, each unit is responsible for a specified job. The MUs units of both client and server are responsible of managing the connection and collaboration session initiation. The user connection process and



collaboration session management do not depend on the used RDS system. In other words, we can say that the MUs of both client and server are common to all RDS systems. The DPU of the client is just responsible for passing the data coming from DSP server to the RDS server/client and vice versa. The DPU during the data passing will not change nor use any item of the passing data it is just passes a stream of bytes without knowing its contents. In other words the DPU is RDS system independent. The Distribution Unit of the DSP platform is divided into two parts the distribution engine and the RDS adapter. The Distribution Engine during the data distributing, analogous to the DPU, will neither change nor use any item of the distributed data. The Distribution Engine distributes a stream of bytes without knowing its contents. Also detecting message process will depend on a general algorithm, named MD-FSM driver. So that, we can say that the Distribution Engine is RDS system independent. The second part of the DU is the RDS adapter. The RDS adapter is the only part of both DSP client and DSP server is RDS system dependent. The RDS adapter has a simple job that is summarized by extracting the setup phase parameters from the messages that were already detected by the distribution unit of the DSP server. So that developing this part will not need large effort and this will satisfy the main objective of the paper.

The data distribution behavior of DSP server enhances the efficiency of the RDS server in terms of bandwidth consumption and processing power. In the tradition direct-connection RDS systems the bandwidth consumption depends on the number of connected RDS clients. Suppose the example in Fig 11. There are four clients joining a collaboration session and the RDS server needs **200KB** to serve one client then the RDS server will consume **800KB**.

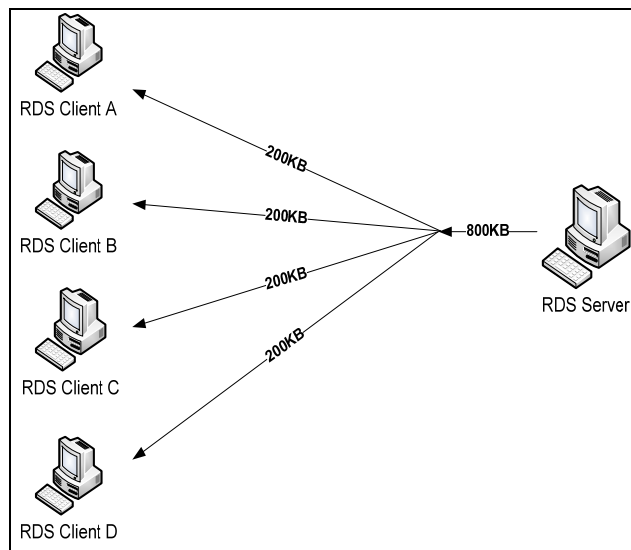


Figure 11. RDS Server's Bandwidth Consumption of traditional RDS system

Every time a new client joining a session it will expand the bandwidth consumption at the RDS server side. The Bandwidth consumption can be calculated from the following equation:

$$B = n.b$$

where,

**B** is the total bandwidth consumption.

**n** Number of joining clients.

**b** is the bandwidth consumption for one user.

In our proposed DSP platform, the RDS server serves only one client and the **DSP** server distributes the exchanging data to other clients so that the bandwidth consumption is optimized from:

$$\text{Number of connecting client} * \text{Bandwidth consumption of one client}$$

To

$$1 * \text{Bandwidth consumption of one client}$$

Fig. 12 shows an RDS server's bandwidth consumption of a traditional RDS system when the bandwidth consumption to serve one client is equal to **200KB**.

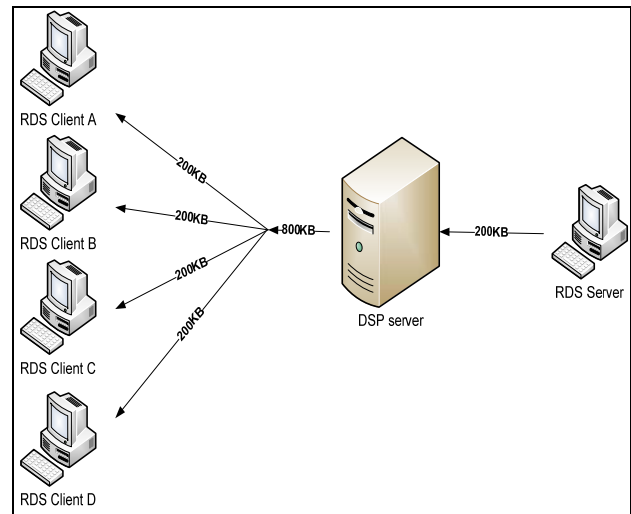


Figure 12. RDS Server's Bandwidth Consumption using the proposed DSP platform

Fig. 13 shows a comparison of the bandwidth consumption at RDS server between the traditional RDS system and the proposed DSP platform when **b** is equal to **200KB**.

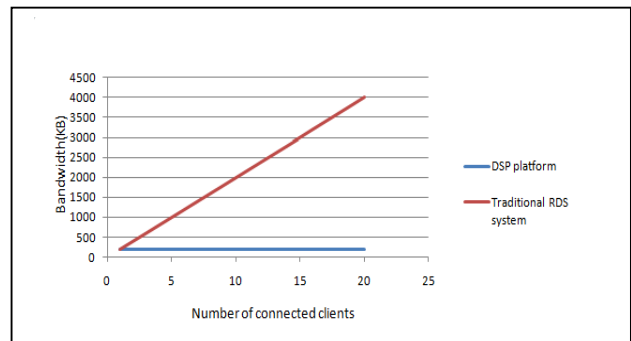


Figure 13 A comparison of RDS Server's bandwidth consumption between the proposed DSP platform and the traditional RDS System

What is applied to bandwidth consumption also can be applied to the processing power at the hosted machine of RDS server. As a result, the processing power increases when the number of connected clients increases. When the RDS server requires high bandwidth and heavy processing in the collaboration session, the hosting of RDS server should be at a machine located within a high bandwidth network and has special properties that make it a suitable of running the RDS server effectively. When adopting the traditional RDS system, there are a few of machines that have ability to host the RDS server. Whereas by adopting the proposed DSP platform, there are many machines that have the ability to host the RDS server.

In the traditional RDS system, it is impossible for the RDS client to connect to the RDS server when the RDS server is located behind Network Address Translation (NAT) unless the RDS server and its clients are located within the same Local Network Area (LAN). To solve this problem the RDS server must be located at a global location. This solution will add another constraint to the machine that has to host the RDS server. This constraint eliminates all the machines behind the NAT from hosting the RDS server. By adopting our proposed DSP platform, only the DSP server is located at global area so that all DSP clients (the DSP clients for RDS clients and the DSP clients for RDS server) can easily connect to its DSP server. This solution of connectivity issues makes most of the current machines to host the RDS server in contrary to the existing RDS system that requires more powerful machines.

## REFERENCES

- [1] Efraim Turban , J.E.A., and Ting Peng Liang, *Decision support systems and Intelligent Systems*. Seventh Edition ed. 2005: Prentice Hall.
- [2] Penichet, V.M.R., et al., *A Classification Method for CSCW Systems*. Electron. Notes Theor. Comput. Sci., 2007. **168**: p. 237-247.
- [3] Clarence, A.E., J.G. Simon, and R. Gail, *Groupware: some issues and experiences*. Commun. ACM, 1991. **34**(1): p. 39-58.
- [4] James, B., R. Mary Beth, and A.S. Clifford, *Flexible collaboration transparency: supporting worker independence in replicated application-sharing systems*. ACM Trans. Comput.-Hum. Interact., 1999. **6**(2): p. 95-132.
- [5] Zhang, L., *Implementing Remote Display on Commodity Operating Systems*, in *Department of Computer Science*. 2006, COLUMBIA UNIVERSITY.
- [6] Richrdson, T., *The RFB Protocol: version 3.8*. 2009, RealVNC Ltd.
- [7] Hui-Chieh, L., et al., *A generic application sharing architecture based on message-oriented middleware platform*. Comput. Stand. Interfaces, 2008. **30**(3): p. 191-199.
- [8] Red Hat, I., *Spice for Newbies*. 2009.
- [9] Forsberg, E., *Reverse-Engineering and Implementation of the RDP 5 Protocol*. 2004, Department of Computer and Information Science, Linköping universitet.
- [10] Lee, D. and M. Yannakakis, *Principles and methods of testing finite state machines-a survey*. Proceedings of the IEEE, 1996. **84**(8): p. 1090-1123.

# Design of Hybrid Ontologies for Mediation System Applied to the E-learning Platform

Otman ABDOUN  
Laboratory LIST,ERIT  
FST – Tangier  
Tangier, Morocco  
otman.abdoun@gmail.com

Jaber EL BOUHDIDI  
Laboratory LIST  
FST - Tangier  
Tangier, Morocco  
jaber.f15@gmail.com

Mohamed GHAILANI  
Laboratory LIST  
FST - Tangier  
Tangier, Morocco  
ghalamed@gmail.com

Abdelhadi FENNAN  
Laboratory LIST  
FST – Tangier  
Tangier, Morocco  
afennan@gmail.com

**Abstract**—This work falls within the scope of E-learning is important for several reasons. First, resources are structured (educational needs) and therefore easier to annotate. Second, there is a curriculum (or Education Plan) that ensures the semantic integration of resources. Third, services are available to the teacher and learner. And finally, post evaluation of knowledge acquired by the learner, to verify the adequacy of resources presented to the learner, and indirectly the appropriateness of teaching strategies implemented to follow up resources and services. First of all, it describes the problems of integrating multiple sources of educational and placed in the ontology integration process, then treated mediation services, and their contribution on an E-learning platform.

**Keywords**- E-learning; Mediation Services; Hybrid Ontologies.

## I. INTRODUCTION

Today, the E-learning platforms and educational information systems use different systems to store and view data. Competition, growth in technology, distribution and evolution of the inevitable decentralization contribute to this plurality. These resources are designed independently of each other, with models and languages that are different, and independent owners. Most of them were not created to be interoperable. To achieve this interoperability, systems integration data are available.

The basic difficulties for integration and heterogeneity of educational resources belong to two concepts: structural and semantic. Using a Mediator Agent ensures translation of responses from different data sources and solves the obstacle of the heterogeneous physical and logical sources or services by providing a uniform access interface. But the semantic heterogeneity remains, even if it requires different sources, to be in a consistent format. One solution involves the use of one or more ontologies as a tool for the integration of educational resources.

## II. INTEGRATION APPROACHES

A data integration system can be characterized by its architecture and integration model. We will distinguish two basic skeletons for data integration.

### A. Mediator Approach

The mediator approach is based on defining mappings for query translation: a request set by the user in terms of global schema is translated into one or more subqueries that are evaluated on resources or services [2]. The answers are ordered and processed to be compatible with the overall pattern and conform to the query posed by the user (Fig. 1).

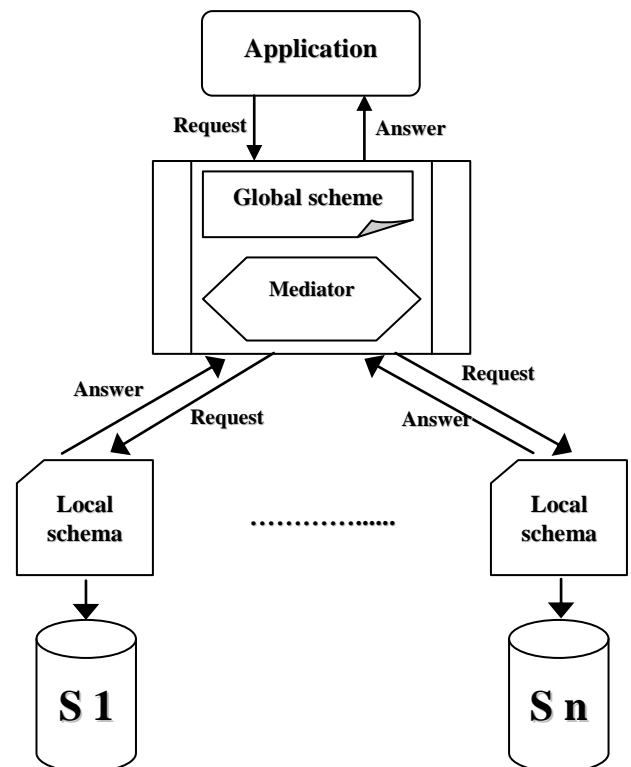


Figure 1. Mediator architecture

### B. Data Warehouse Approach

The warehouse approach applies the principle of the views and integrates data sources in accordance with the overall patterns [7]. The result is a data warehouse that can be directly examined through a suitable language (Fig. 2).

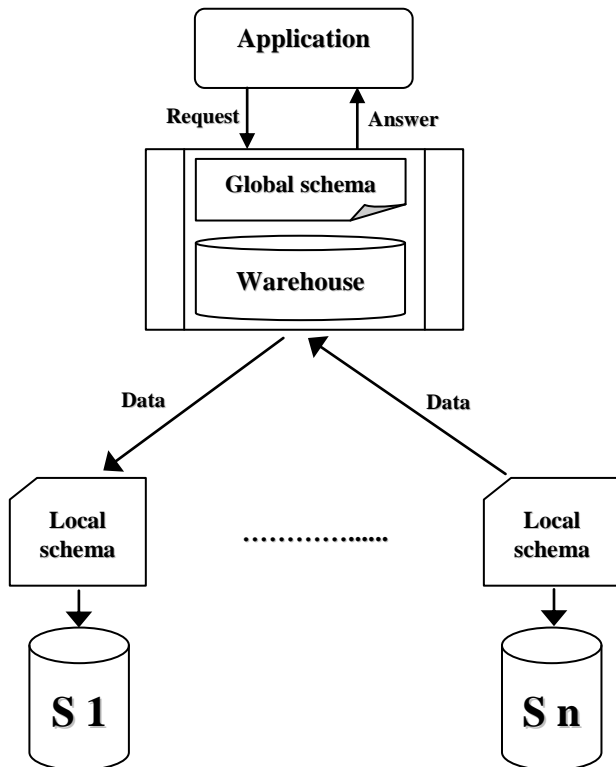


Figure 2. Mediator Warehouse

## III. DATA INTEGRATION BASED ON ONTOLOGY

To achieve semantic interoperability in heterogeneous information, system requires that the semantics of information exchanged is understood throughout the system.

*The Ontology gives the names and descriptions of entities in a specific field by using the attributes that represent the relationship between these entities.*

There are many advantages in the use of ontologies for data integration. The ontology provides a rich vocabulary and predefined concept that interfaces stable access to databases, and is independent of database schemas. Knowledge represented by the ontology is sufficiently complete to support the appropriate translation of all sources of information [1]. The ontology supports compliance management and identification of conflicting data.

The use of ontologies for the interpretation of implicit and hidden knowledge is one possible approach to overcome the problem of semantic heterogeneity. Many approaches to integration based on ontologies have been developed to achieve interoperability.

In almost all approaches to integration based on ontologies, they are used for the explicit description of the semantics of information sources. But how to use these ontologies can be different. Three different directions are identified as follows:

### A. Approach with a Single Ontology

*The approach with a single ontology* which uses a global ontology that provides a shared vocabulary for the specification of the semantics of data sources (Fig. 3). All data sources are linked to a global ontology. This can also be a combination of specialized ontologies. We can apply this approach to integration problems where all information sources to integrate provide almost the same view on a domain.

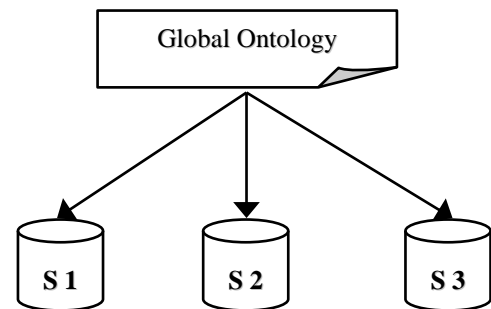


Figure 3. Approach with a single ontology

This approach has a major drawback when adding or removing data sources. Indeed, the conceptualization of the domain represented in the ontology may require changes. This led to the development of approaches with multiple ontologies.

### B. Approach Based on Multiple Ontologies

In the approach with several ontologies, each source is described by its own ontology (Fig. 4). The advantage of this approach is that ontology has no need for commitment to a common minimum global ontology. Each source of ontology can be developed without the need to meet or find other sources and ontologies. This architecture can significantly simplify the task of integrating and supporting the change (adding and removing sources). However, the lack of common vocabulary makes it difficult to compare between different source ontologies.

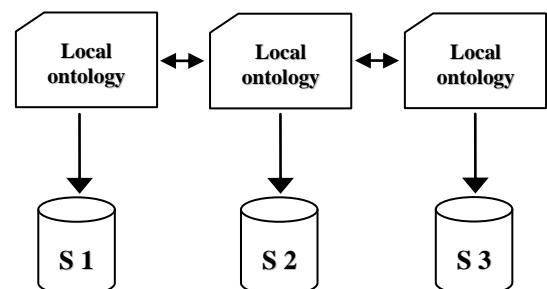


Figure 4. Approach based on multiple ontologies

To overcome this problem, an additional, formal representation defining the mapping between ontologies is necessary. The mapping between ontologies semantically identifies the correspondence of terms of different ontologies.

### C. Hybrid Approach

To overcome the drawbacks of the first two approaches, **hybrid approaches** have been developed (Fig. 5). This approach describes the semantics of each source by its own ontology as with the approach to multiple ontologies. But to make the local ontologies comparable to each other, they are built from a global shared vocabulary.

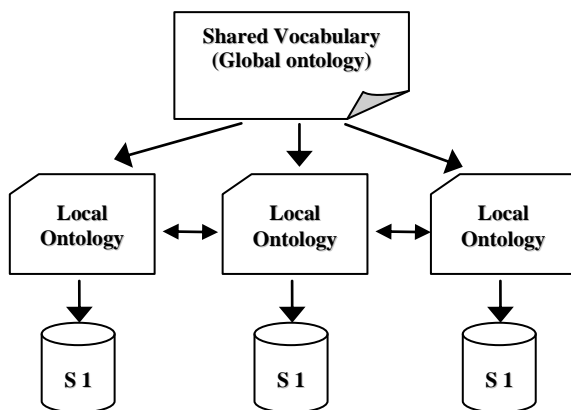


Figure 5. Hybrid approach for the description of data sources

It contains the basic terms is an area that can be combined with local ontologies to describe a more complex semantics. Sometimes the shared vocabulary can be an ontology.

The advantage of the hybrid approach is the fact that new sources can easily be added without the need for change. Also, this approach supports the acquisition and development of ontologies. But the major drawback of hybrid approaches is that existing ontologies cannot easily be reused, but must be rebuilt.

The state of the art in data integration architecture showed that the hybrid approach allows for greater scalability and extension. Indeed, this architecture allows adding new sources to ensure certain independence.

The mediation system must manage the independence of data sources and their distribution. In addition, the system must manage the interaction between the global ontology and local ontologies in creating queries.

## IV. MEDIATION ARCHITECTURE ADAPTED FOR A PLATFORM FOR DISTANCE LEARNING

E-learning application is online through the use of the Web. Given the diversity and the exponential growth of learning resources used in a training type E-learning, it is increasingly difficult to find relevant teaching materials [8]. E-learning application is sharing the same problem of relevance with the Web when learners want to access knowledge at their disposal.

### A. The Modeling of a Mediation System Based on Ontologies for E-Learning Platform

We agree to use a mediation system based on ontologies. Local and global ontologies provide a common set of terms that can be applied to any resource, which allows organizations to describe and search their resources [11].

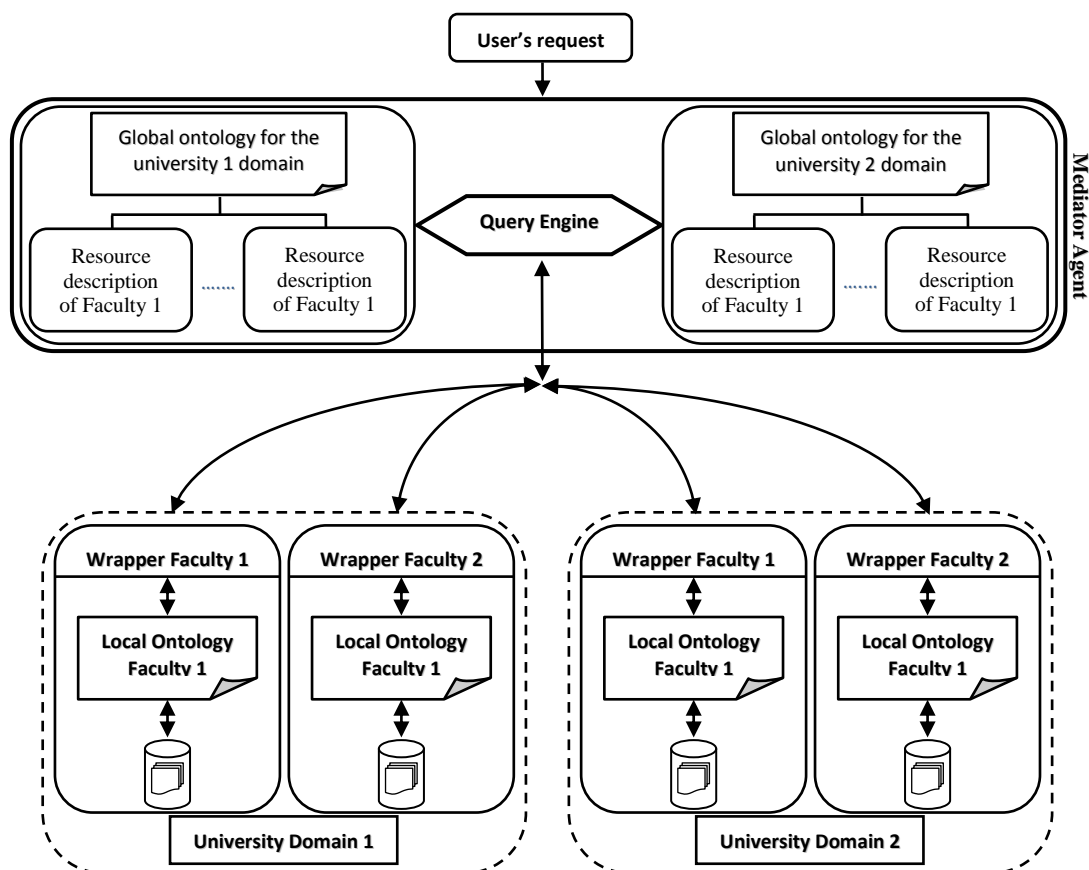


Figure 6. Proposed mediation architecture

Facing the rapidly changing media and communication in the field of machine learning, a critical mass of educational resources is produced across many universities. Reuse of learning objects, thus produced locally, is low and even non-existent among universities. This can be explained, in particular by the lack of knowledge on the part of teachers on existing resources. We propose a model for integrating data sources that aims to overcome this deficit. This is a mediation system resources distributed based on the description by the local ontology and a global ontology, this scenario is described in the above (Fig. 6).

The local ontology contains the description of sources. And a global ontology represents a domain ontology learning resources (Fig. 7).

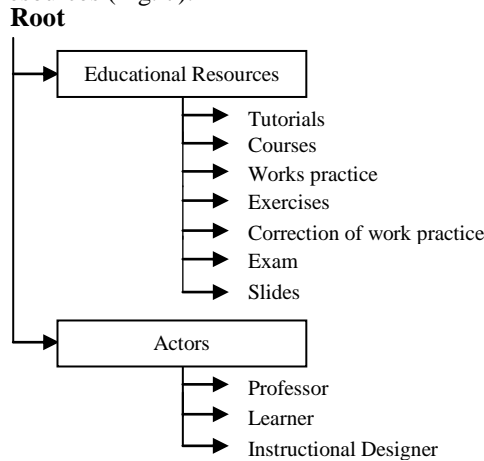


Figure 7. Ontology for a local educational resource

We assume in our approach that each university has its own ontology, a description of learning resources and their semantic description. Domain ontologies and descriptors of each source are, then, used to build knowledge on distributed objects, like a catalog accessible to all.

### B. Scenario for Integration of Educational Resources

An actor system (Student, Professor, ...) calls for a teaching aid while specifying a number of criteria that describe it (Speciality, Material, Level, Type of document format). The mediator agent (MA) when he receives the request, he dissected to extract the specialty that will be sought in the global ontology to index academic institutions responding to the request. Then, AM makes a request for it includes in addition to the Specialty Material Level, Type and Format document and send it to agents Wrapper institutions indexed. Each wrapper agent sends the request received after his translation to make it understandable by the agent resources. The latter consults the ontology to search for local media requested, and sends the result to find the agent wrapper, which in turn will translate and send the response to the mediator agent. AM generates a page for the user, indexing all media found. The figure 8 describes the sequence diagram for the processing steps of the integration of educational resources.

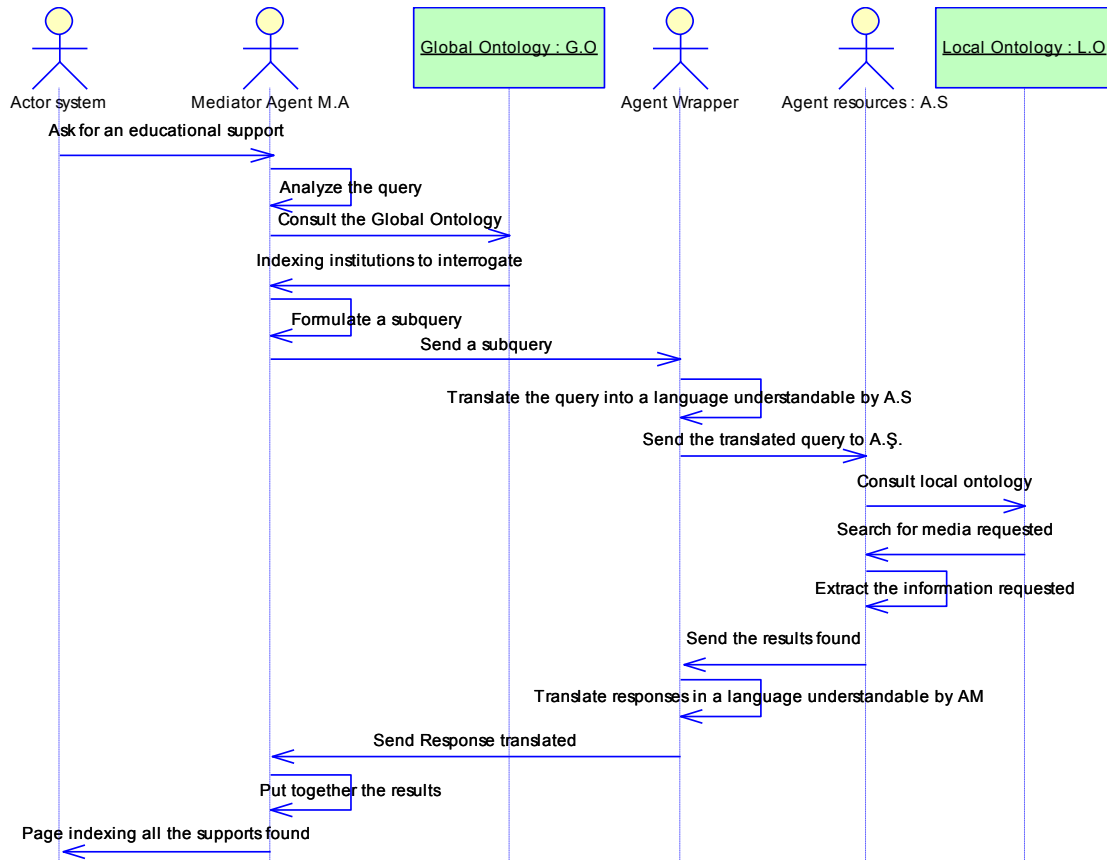


Figure 8. A Sequence Diagram for the processing steps for integration of educational resources



## CONCLUSION

In this paper work, we presented a new mediation architecture with an objective to build up an environment for integrating various educational sources / services, achieve interoperability and heterogeneity between these sources, and consult and look for educational materials.

The concepts of data integration and services based on ontologies, and several approaches to use them in the integration of data sources, were presented. We showed the provision of mediation systems to E-learning, referring to some mediation projects, which have been made based on ontologies.

## REFERENCES

- [1] J. El Bouhdidi, M. Ghailani, O. Abdoun, A. Fennan, A New Approach based on a Multi-ontologies and Multi-agents System to Generate Customized Learning Paths in an E-Learning Platform, International Journal of Computer Applications, 2010.
- [2] Kalpana Nigam, Monica Mehrotra, Ontology Mediation with Machine Learning, IJCSIC, 2010
- [3] J. Euzenat and P. Shvaiko. Ontology Matching. Springer, 2007.
- [4] A. Gangemi. Ontology design patterns for semantic Web content. In Y. Gil, E. Motta, V. Richard Benjamins, and M.A. Musen, editors, International Semantic Web Conference, volume 3729 of Lecture Notes in Computer Science, pages 262–276. Springer, 2005.
- [5] A. Coulet, M. Smaïl-Tabbone, A. Napoli, and M.D. Devignes. Role Assertion Analysis : a proposed method for ontology refinement through assertion learning. In Proceedings of the Fourth Starting AI Researchers' Symposium (STAIRS 2008), pages 47–58. IOS Press, 2008.
- [6] Mohand-Saïd Hacid et Chantal Reynaud : L'intégration de sources de données, Thèse, LIRIS, UFR Informatique Université Claude Bernard Lyon 1, 2009
- [7] T. Adams, J. Dullea, P. Clark, S. Sripada, T. Barrett. Semantic Integration of Heterogeneous Information Sources Using a Knowledge-Based System. In Proc 5th IntConf on CS and Informatics (CS&I'2000), 2005, pages
- [8] S. Barlow. Data Integration. University of Passau. July 24, 2006,
- [9] Abel MH., Dieng-Kuntz R., Hérin D., Lenne D., Moulin C., Pompidor P. Langages pour le Web Sémantique et pour le E-Learning. Journée thématique AFIA : Web sémantique pour le e-Learning. Plateforme AFIA. Nice, 30 mai. pp. 97-122, 2005.

- [10] S. Cohen-Boulakia, O. Biton, S.B. Davidson, and C. Froidevaux. BioGuideSRS : querying multiple sources with a user-centric perspective. Bioinformatics, 23(10) :1301–1303, 2007.
- [11] Y. An, J. Mylopoulos, and A. Borgida. Building semantic mappings from databases to ontologies. In AAAI. AAAI Press, 2006.
- [12] Buffa, M., Dehors, S., Faron-Zucker, C., Sander, P. (2005), Vers une approche sémantique dans la conception d'un système d'apprentissage. Revue du projet TRIAL-SOLUTION, Plate forme AFIA Nice,
- [13] B. Berendt, A. Hotho, and G. Stumme, editors. Proceedings of the Workshop on Semantic Web Mining (SWM'02 at ECML/PKDD'02), Helsinki, Finland, August 2002.
- [14] Peter Brusilovsky, C.K., Demetrios Sampson, Layered evaluation of adaptive learning systems. International Journal of Continuing Engineering Education and Lifelong Learning, 2004.

## AUTHORS PROFILE

**Otman ABDOUN.** Born in KSAR EL KEBIR, Morocco. He received the “Master of Sciences and Technologies” in Computer Engineering and the “Magister in Computer Sciences Systems and Networks” degree from the University of Abdelmalek Essaadi, Tangier, Morocco. He is currently working toward the Ph.D. degree with the Computing and Telecommunications Research Group at Abdelmalek Essaadi University, Tangier, Morocco. His research interests include Mediation and semantic web in E-Learning.

**Jaber EL BOUHDIDI.** Born in Chefchaouen, Morocco. He received the “Master of Sciences and Technologies” in Computer Engineering and the “Magister in Computer Sciences Systems and Networks” degree from the University of Abdelmalek Essaadi, Tangier, Morocco. He is currently working toward the Ph.D. degree with the Computing and Telecommunications Research Group at Abdelmalek Essaadi University, Tangier, Morocco. His research interests include semantic web and Multi-agents Systems in E-Learning.

**Mohamed GHAILANI.** Born in Larache, Morocco. He received the “Master of Sciences and Technologies” in Computer Engineering and the “Magister in Computer Sciences Systems and Networks” degree from the University of Abdelmalek Essaadi, Tangier, Morocco. He is currently working toward the Ph.D. degree with the Computing and Telecommunications Research Group at Abdelmalek Essaadi University, Tangier, Morocco. His research interests include web semantic, ontology in E-Learning.

**Abdelhadi FENNAN.** is PhD doctor and professor of computer science at Faculty of Sciences and technology of Tangier in Morocco. He is part of many boards of international journals and international conferences. He has published several articles on E-learning.

# Extracting Membership Functions Using ACS Method via Multiple Minimum Supports

Ehsan Vejdani Mahmoudi

Islamic Azad University, Mashhad  
Branch, Young Researchers Club,  
Mashhad, Iran  
e.vejdani@mshdiau.ac.ir

Masood Niazi Torshiz

Department of computer  
Engineering, Islamic Azad  
University - Mashhad Branch,  
Mashhad, Iran  
niazi@mshdiau.ac.ir

Mehrdad Jalali

Department of computer  
Engineering, Islamic Azad  
University - Mashhad Branch,  
Mashhad, Iran  
jalali@mshdiau.ac.ir

**Abstract**— Ant Colony Systems (ACS) have been successfully applied to different optimization issues in recent years. However, only few works have been done by employing ACS method to data mining. This paper addresses the lack of investigations on this study by proposing an ACS-based algorithm to extract membership functions in fuzzy data mining. In this paper, the membership functions were encoded into binary bits, and then they have given to the ACS method to discover the optimum set of membership functions. By considering this approach, a comprehensive exploration can be executed to implement the system automation. Therefore, it is a new frontier, since the proposed model does not require any user-specified threshold of minimum support. Hence, we evaluated our approach experimentally and could reveal this approach by significant improving of membership functions.

**Keywords**- fuzzy data mining; multiple minimum supports; association rule; membership functions; ant colony system.

## I. INTRODUCTION

Recently, the fuzzy set theory has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning [1]. As to fuzzy data mining, Hong and Kuo proposed a mining approach that integrated fuzzy-set concepts with the Apriori mining algorithm [2].

ACO is a branch of a larger field referred to as Swarm Intelligence (SI). SI is the property of a system whereby the collective behaviors of simple agents interacting locally with their environment cause coherent functional global patterns to emerge [3]. It is the behavioral simulation of social insects such as bees, ants, wasps and termites. This behavioral simulation came about for many reasons—optimization of systems and learning about self-organization are two of many reasons why scientists are interested in simulating these insects. More specifically, ACO simulates the collective foraging habits of ants—ants venturing out for food, and bringing their discovered food back to the nest. Ants have poor vision and poor communication skills, and a single ant faces a poor probability of longevity. However, a large group, or swarm, of ants can collectively perform complex tasks with proven effectiveness, such as gathering food, sorting corpses or performing division of labor [4]. They are a heuristic

approach inspired from the behavior of social insects. Ants deposit their chemical trails called “Pheromone” on the ground for communicating with others. According to the pheromone, ants can find the shortest path between the source and the destination. Recently, Ant Colony Systems (ACS) has been successfully applied to several difficult NP-hard problems, such as the quadratic assignment [5], communication strategies [6], production sequencing problem [7]. Job Schedule Problem (JSP) [8], the traveling salesman problems [9], [10], Vehicle Routing Problems (VRP) [11], etc.

Basically, fuzzy mining algorithms first used membership functions to transform each quantitative value into a fuzzy set in linguistic terms and then used a fuzzy mining process to find fuzzy association rules. Items have their own characteristics, different minimum supports specified for different items. Han, Wang, Lu, and Tzvetkov [12] have pointed out that setting the minimum support is quite subtle, which can hinder the widespread applications of these algorithms. Our own experiences of mining transaction databases also tell us that the setting is by no means an easy task. Therefore, our approach proposed method for computing minimum supports for each item in database with own features. This approach leads to effectiveness, efficiency for global search and system automation, because our model does not require the user specified threshold of minimum support.

Numerical experiments on the proposed algorithm are also performed to show its effectiveness. The remaining parts of the paper are organized as follows. Section II presents An ACS-based mining framework. The proposed algorithm based on the above framework is described in Section III. Numerical simulations are shown in Section IV. Conclusions are given in Section V.

## II. THE ACS-BASED FUZZY MINING FRAMEWORK

In this section, the ACS based fuzzy mining framework [13] is shown in Fig. 1 where each item has its own membership function set. These membership function sets are then fed into the ant colony system to search for the final proper sets. When the termination condition is reached, the best membership function set (with the highest fitness value) can then be used to mine fuzzy association rules from a database.

The proposed framework modified the ACS-based Framework for Fuzzy Data Mining in [13]. The framework is divided into two phases. The first phase searches for an appropriate set of membership functions for the items by the ACS mining algorithm. Having searched for the solutions in the first phase, we use the best membership functions for fuzzy data mining in the second phase.

The ACS algorithm plays an important role in extracting the membership functions. In the past, Parpinelli et al. proposed the AntMiner to discover association rules [14]. They worked on categorical attributes and discrete values. They proved that the ACS algorithm performed well on handling discrete values in a solution space. In this work, we assume the parameters of membership functions as discrete values and thus try to use the ACS algorithm to find them. We transform the extraction of membership functions into a route-search problem. A route then represents a possible set of membership functions. The artificial ants, which refer to virtual ants that are used to solve this problem, can then be used to find a nearly optimal solution.

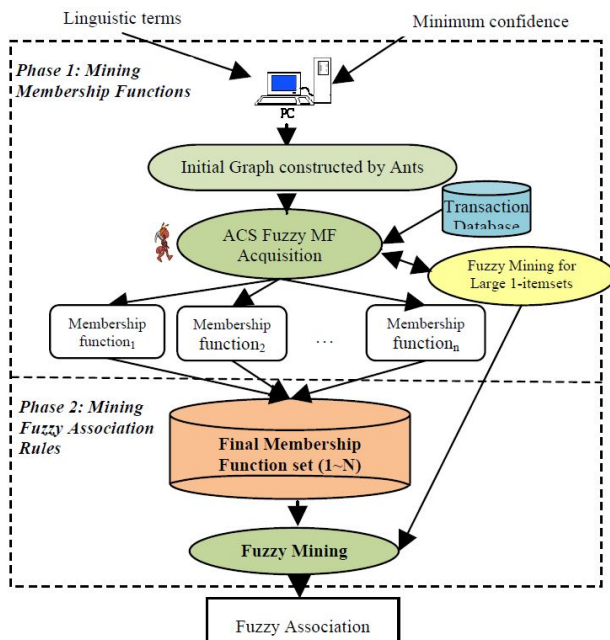


Figure 1. The ACS-based framework for fuzzy data mining

### III. THE ACS\_BASED FUZZY DATA MINING ALGORITHM

#### A. Initializations

As revealing membership functions of all items result in a long code, we will encode the membership function of each item into a binary code. We use the coding algorithm which was represented in [13]. Furthermore, we utilize some rules called State transition rule, Pheromone updating rule, Local updating rule, Global updating rule which were defined in [15].

In this work, each item will have a set of isosceles-triangular membership functions. The membership function stands for the linguistic terms such as low, middle, high. Transforming these quantitative values into linguistic terms requires a feasible population of database. Therefore, we need to initialize and update a population during the evolution

process. In this work, we use the fitness function proposed by Chen et al. [16] to obtain a good set of membership functions.

#### B. ACS-based fuzzy data mining algorithm

Although, the proposed algorithm as considered in [15] and [16] concerns one constant minimum support for all items, we applied the determined minimum support for each item. As a matter of fact, in real world applications such as work on transactional data of chain stores, the items have different quantities. Hence, using different minimum supports for each item in order to extracting membership functions is an efficient idea. However, the previous ones that user specified minimum supports, the new approach proposes the minimum supports are achieved by a preprocessing on all items. On the other hand, minimum support for each item is automatically set as a value correspond with the quantity of the item. We considered a method for computing minimum support for each item with its characteristics in databases. There are significant criteria for computing minimum support like, the number that each item happened in database and sum of values for each item in database. For example, suppose the number that item A happened in database is 10 and sum values is 20 and also the number that item B happened in database is 2 and sum values is 20. Clearly in mining process item A valuable than item B. We computing minimum support for item B until this item can't satisfying minimum support. As mentioned above, we suggested in (1) as below:

$$\min\_Sup(I_i) = \frac{\sum_{j=1}^n S_j}{T * N * P} \quad (1)$$

Let  $I = \{i_1, i_2, \dots, i_m\}$  be a set of items and  $D = \{t_1, t_2, \dots, t_n\}$  be a set of transactions.  $N$  is total number of transaction data.  $T$  is the number that each item happened in database.  $S_j$  is sum values of an item in database  $D$ .  $P$  is constant digit with respect to the interval  $[0, 1]$ .

In addition, as we investigated the parameters defined in [13], the following parameters performed: The number of artificial ants, the minimum pheromone ratio of an ant, the evaporation ratio of pheromone, the local updating ratio, and the global updating ratio. The proposed ACS-based algorithm for mining membership functions and fuzzy association rules are given as follow.

#### INPUT :

- quantitative transaction data,
- a set of  $m$  items, which is with  $l$  predefined linguistic terms,
- a maximum number of iterations  $G$ ,
- $P$  is constant digit with respect to the interval  $[0, 1]$ .

OUTPUT: An appropriate set of membership functions for all items in fuzzy data mining.

step 1) Let  $p = 1$ , where  $p$  is used to keep the identity number of the items to be processed.

step 2) Let the multi-stage graph for the fuzzy mining problem be  $(N, E)$ , where  $N$  is the set of nodes and  $E$  is the set of edges. Also denote the  $j$ -node in the  $i$ -th stage as  $N_{ij}$ , and

the edge from  $N_{ij}$  to  $N_{(i+1)k}$  as  $N_{ijk}$ . Initially set the pheromone on every edge  $N_{ijk}$  as 0.5.

step 3) Let the initial generation  $g = 1$ .

step 4) Sets up the complete route for each artificial ant  $Ant_q$  by the following sub steps.

a) Selects the edges from start to end according to the state transaction rule.

b) Update the pheromone of the edges passed through by  $Ant_q$  according to the local updating rule.

step 5) Evaluate the fitness value of the solution (membership functions) obtained by each artificial ant according to the following sub steps.

a) For each transaction datum  $D_i$ ,  $i = 1$  to  $n$ , transfer its quantitative value  $v_p$  for item  $I_p$  into a fuzzy set  $f_p$  according to the membership functions obtained from the ant in (2). That is,  $f_p$  is represented as :

$$\frac{f_{p1}}{Region_{p1}} + \frac{f_{p2}}{Region_{p2}} + \dots + \frac{f_{pi}}{Region_{pk}} + \dots + \frac{f_{pl}}{Region_{pl}} \quad (2)$$

Where  $Region_{pk}$  is the  $k$ -th fuzzy term of item  $I_p$ ,  $f_{pk}$  is  $v_p$ 's fuzzy membership value in the region, and  $l$  is the number of fuzzy membership functions.

b) The scalar cardinality of each region in the transactions is calculated in (3):

$$count_{pk} = \sum_{i=1}^n f_{pk}^{(i)} \quad (3)$$

Where  $f_{pk}^{(i)}$  is the fuzzy membership value of region  $R_{pk}$  from the  $i$ -th datum.

c) Check for each  $R_{pk}$  whether its  $count_{pk} / n$  is larger than or equal to the minimum support threshold  $\alpha$ . If  $R_{pk}$  satisfies the above condition, put it in the set of large 1-itemsets ( $L_1$ ).

d) Calculate the fitness value of the solution from the ant by dividing the number of large itemsets in  $L_1$  over the suitability. That is Equation(4),

$$fitness = \frac{|L_1|}{suitability} \quad (4)$$

step 6) Once all the artificial ants find their entire routes, the one holding the highest fitness value will be used to update the pheromone according to the global updating rule.

step 7) If the generation  $g$  is equal to  $G$ , output the current best set of membership functions of item  $I$  for fuzzy data mining; otherwise,  $g = g + 1$  and go to  $s.4$ .

step 8) If  $p \neq m$ , set  $p = p + 1$  and go to Step 2 for another item; otherwise, stop the algorithm.

The final set of membership functions output in step 7 and the 1-itemsets obtained are then used to mine fuzzy association rules from the given database.

#### IV. NUMERICAL SIMULATION

We experimentally evaluated our approach to expose the performance of the proposed algorithm. The experiments were implemented in C/C++ on a computer with Intel Core(TM) 2 Duo Processor 2.66GHz and 4 GB main memory, running the Microsoft Windows 7 operating system. We used two datasets to present results: Dataset [13] with a total of 64 items and 10,000 transactions. In addition, a real dataset called FOODMART from an anonymous chain store was used in the experiments [17]. The FOODMART dataset contained quantitative transactions about the products sold in the chain store. There were totally 21,556 transactions with 1600 items in the dataset Used in the experiments. The initial count of ants was set at 10. The parameters in the ACS algorithm were set as follows: the initial ratio of pheromone was 0.05, the minimum pheromone of ants was 0.2, the evaporation ration was 0.9, the local updating ratio was 0.1 and the global updating ratio was 0.9, minimum support for FOODMART dataset was set to 0.0015 and for dataset [13] was set to 0.04. We considered the value of constant  $P$ , as mentioned in (1) for FOODMART dataset equal to 0.05 and for dataset [13] equal to 0.02.

The average fitness values of the artificial ants along with different numbers of generations for two datasets are shown in Fig. 2 and Fig. 3.

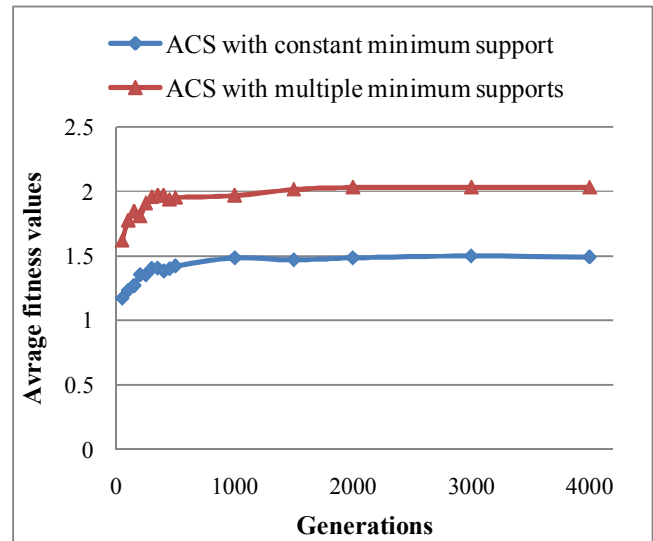


Figure 2. The average fitness values along with different numbers of generations with dataset [13]

It can be vividly seen from Fig. 2 and Fig. 3 that in our approach, the average fitness values increased by an offset compared with the previous one. Thus, became stable within less number of generations. In addition, we used smaller numbers for generation with the aim of comparing the difference between our model and the existing one that has static constant minimum support in Fig. 4. It is obviously represents that our model achieved the best fitness at 300 numbers of generations, whereas the existing one reached its best fitness at 500 numbers of generations.

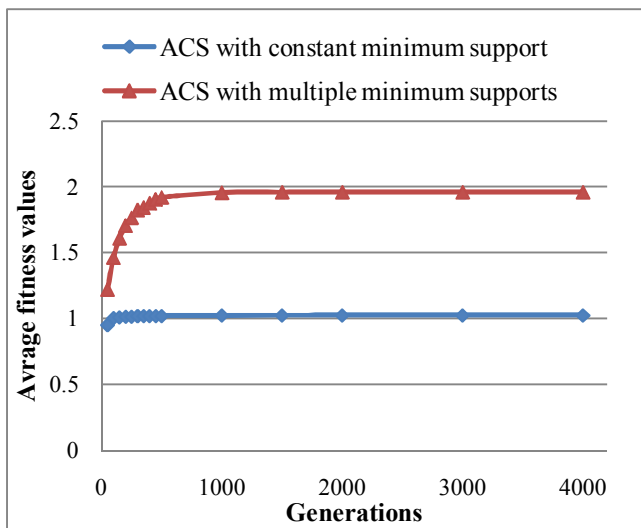


Figure 3. The average fitness values along with different numbers of generations with FOODMART dataset

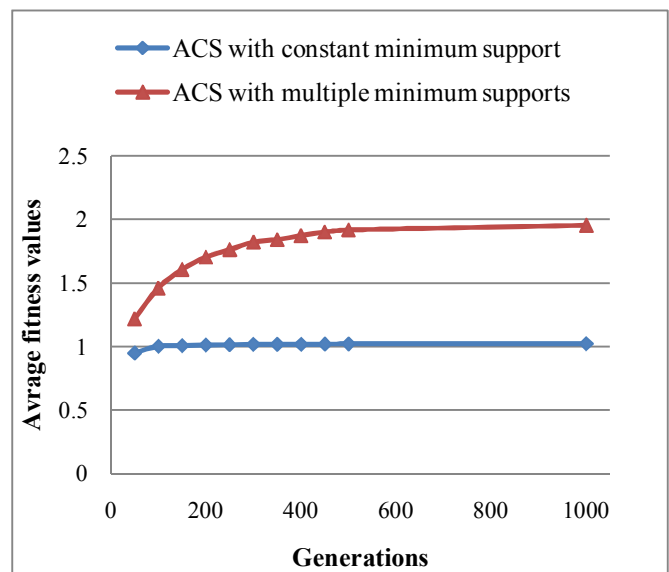


Figure 5. The average fitness values along with different numbers of generations with FOODMART dataset (in smaller scale)

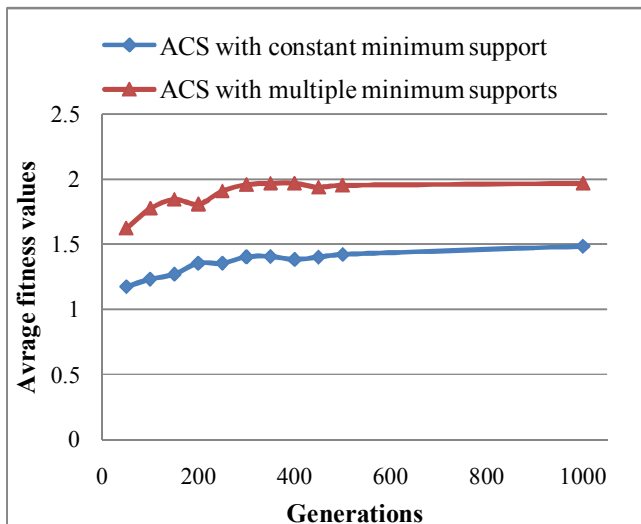


Figure 4. The average fitness values along with different numbers of generations with dataset [13] (in smaller scale)

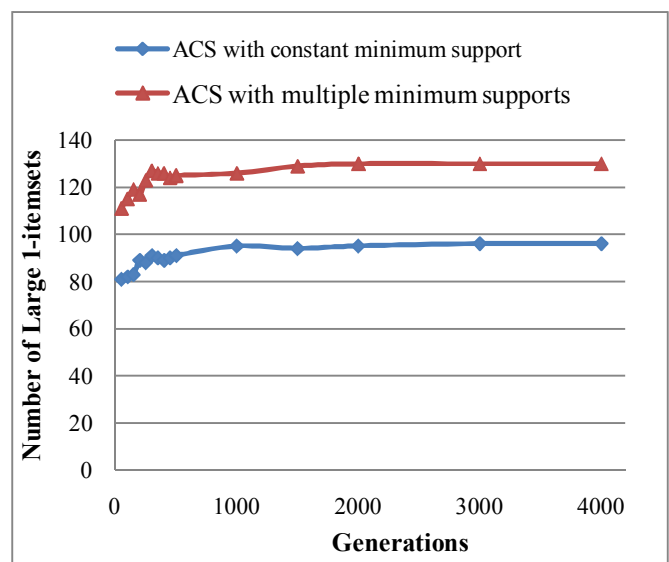


Figure 6. The numbers of large 1-itemsets along with different numbers of generations with dataset [13]

As shown in Fig.5, the result of executing ACS algorithm with multiple minimum supports on FOODMART dataset is much better than ACS algorithm with constant minimum support since it has higher average of fitness values. The number of items in FOODMART dataset is too many. Therefore artificial ants have been through difficulty for optimizing membership functions. Meanwhile, ACS algorithm with multiple minimum supports could easily pass this test, and extracting membership functions with high average of fitness values.

The numbers of large 1-itemsets along with different generations are shown in Fig. 6. The curve of the existing method stabilized after about three thousand generations while the curve of our approach remained constant after one thousand generations. Besides, the number of large 1-itemsets of our approach is clearly much higher.

Fig.7 illustrates the numbers of large 1-itemsets along with different generations for FOODMART dataset. The proposed ACS algorithm could increase large 1-itemsets in interval 50 to 500 generations, and stabilize after about 500 generations while the existing method with increasing generation had no changes, since the existing algorithm cannot work with FOODMART dataset which have a lot of items.

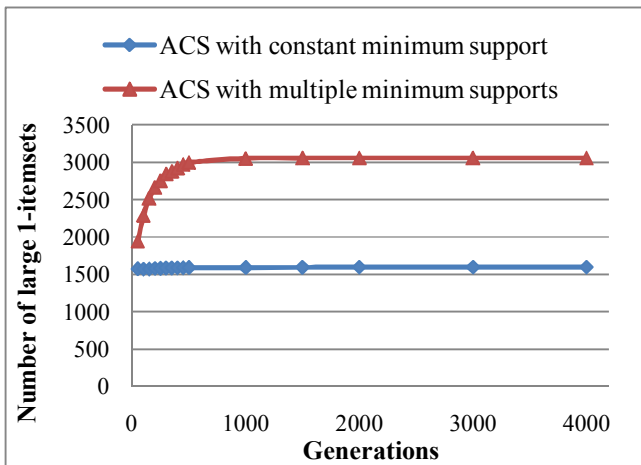


Figure 7. The numbers of large 1-itemsets along with different numbers of generations with FOODMART dataset

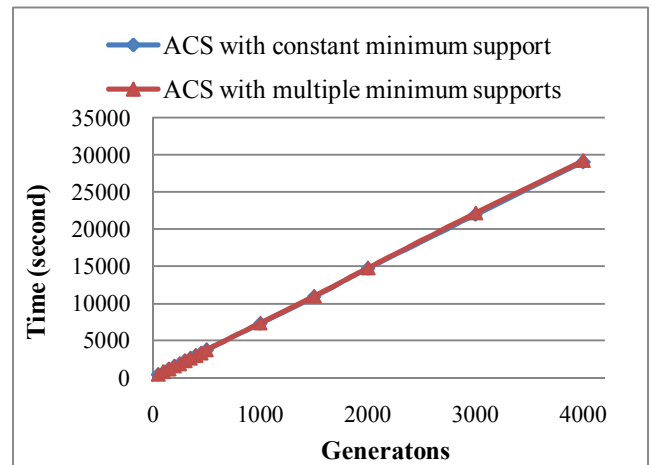


Figure 9. The execution time of the ACS mining algorithm with FOODMART dataset

Fig. 8 and Fig. 9 reveal the execution time of the ACS algorithms for different numbers of generations. Although, execution time increased along with the generations within both line graphs. Therefore, our approach represents the same execution time for smaller number of generations, but increases for high number of generations, slightly.

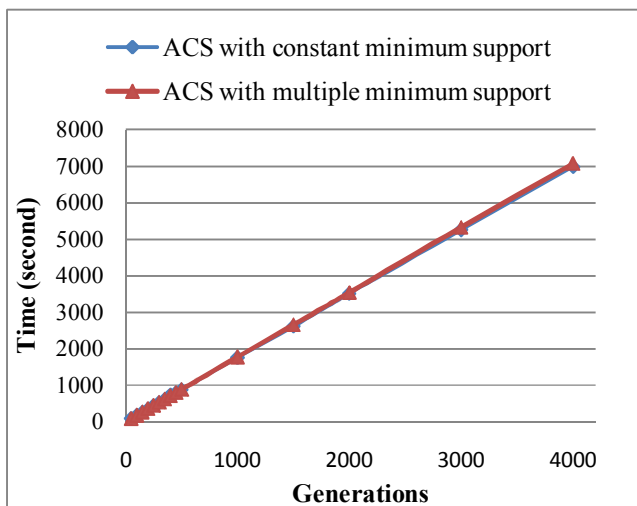


Figure 8. The execution time of the ACS mining algorithm with dataset [13]

In the following study, we expressed the ACS algorithms efficiency with scalability test on two datasets. The generation parameter among execution of algorithms is considered with constant value of 500. The average of fitness values of the artificial ants along with different size of dataset [13] is shown in Fig. 10. By increasing the size of dataset, the accurate membership functions are extracted, and the artificial ant can learn more and find proper solutions. While in existing algorithm with increasing the size of dataset has no changes. Fig. 11 which is executed on FOODMART dataset, is as the same as Fig.10 mentioned before.

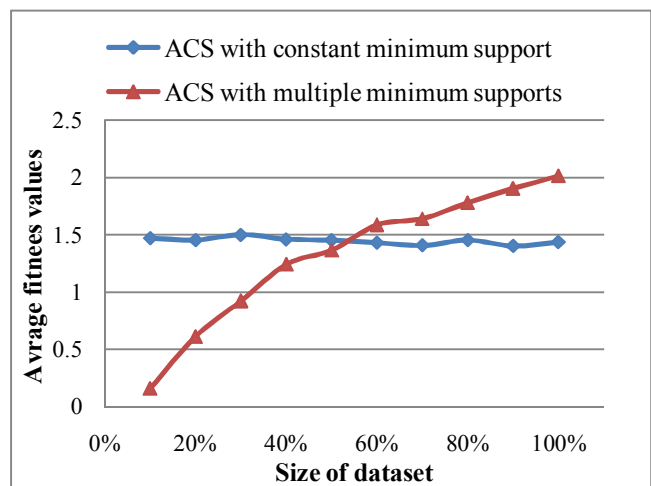


Figure 10. The average fitness values along with different size of dataset [13]

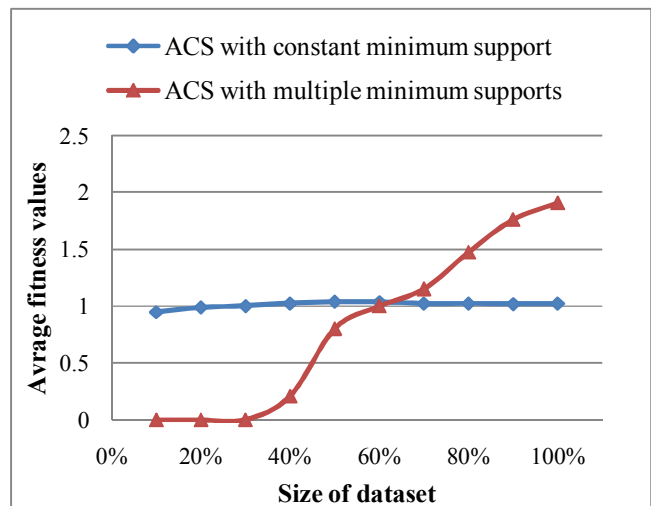


Figure 11. The average fitness values along with different size of FOODMART dataset



The large 1-itemsets of the artificial ants along with different size of datasets is shown in Fig. 12 and Fig. 13. By increasing the size of dataset, the number of large 1-itemsets increased as well. However, at the first existing algorithm had high values. Nevertheless, the ACS with constant minimum support had remained steady.

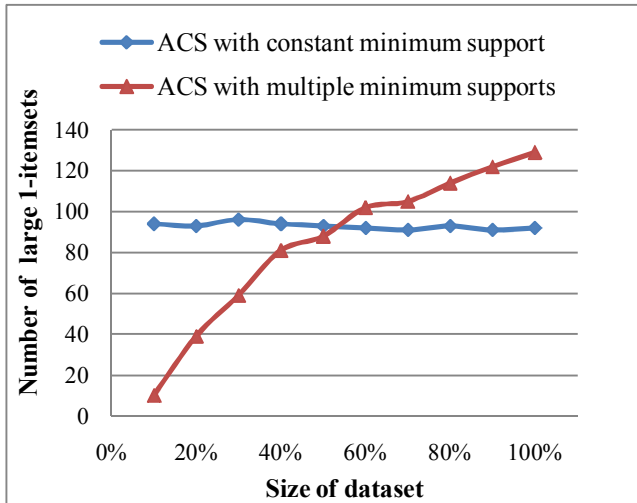


Figure 12. The numbers of large 1-itemsets along with different size of dataset [13]

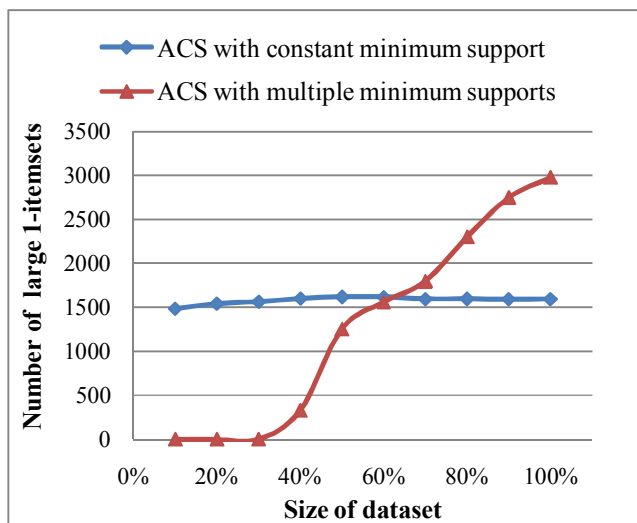


Figure 13. The numbers of large 1-itemsets along with different size of FOODMART dataset

Fig. 14 and Fig. 15 reveal the execution time of the ACS algorithms for different size of datasets. As can be observed in Fig. 14 and 15, the execution time of both algorithms is nearly equal. It therefore proves that proposed algorithm does not increase the execution time as well as improving efficiency encourages us to employ the proposed algorithm for extracting membership functions.

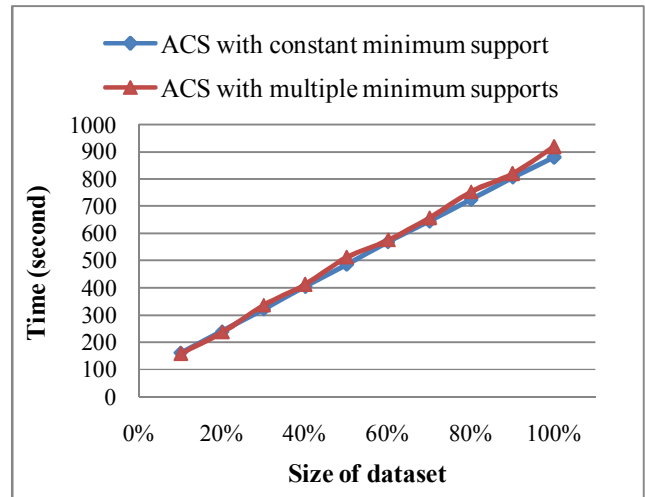


Figure 14. The execution time of the ACS mining algorithm along with different size of dataset [13]

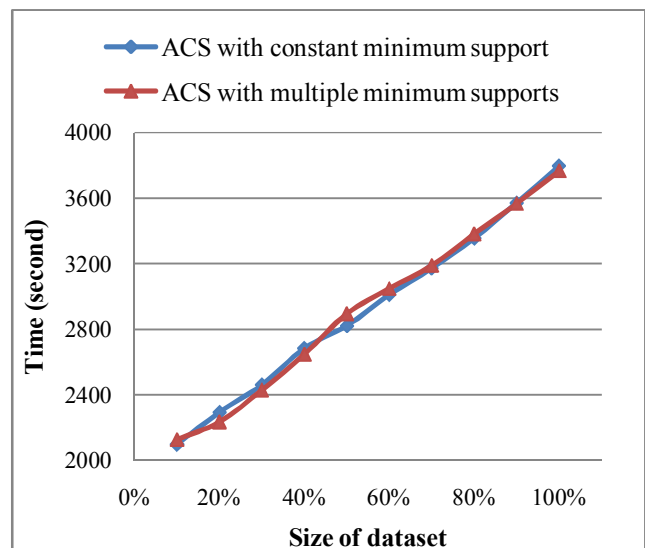


Figure 15. The execution time of the ACS mining algorithm along with different size of FOODMART dataset

## V. CONCLUSIONS

In this paper, we could seek for the issues of applying the ACS algorithm to extract membership functions for fuzzy data mining and have proposed an algorithm to address this aim. As a matter of fact, in this approach we could deliver two benefits including the usage of multiple minimum supports, and system automation. On the other hand, computation results illustrated our work can be given as an alternative for effective association rule mining.

Meanwhile, the most significant difference between our algorithm and older ACS algorithms to extract membership functions concerns the independency of minimum support threshold. The experimental results of this new approach encouraged us to improve the system and utilize this strategy in real world applications, magnificently.

## REFERENCES

- [1] Kandel, "Fuzzy expert systems," *CRC Press*, pp. 8- 19, 1992.
- [2] T. P. Hong, *et al.*, "Trade-off between time complexity and number of rules for fuzzy mining from quantitative data," *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, vol. 9, pp. 587-604, 2001.
- [3] E. Bonabeau, *et al.* (1999). *Swarm intelligence from natural to artificial isystems*. Available: [http://link.library.utoronto.ca/eir/EIRdetail.cfm?Resources\\_ID=432271&T=F](http://link.library.utoronto.ca/eir/EIRdetail.cfm?Resources_ID=432271&T=F)
- [4] W. B. Tao, *et al.*, "Object segmentation using ant colony optimization algorithm and fuzzy entropy," *Pattern Recognition Letters*, vol. 28, pp. 788-796, May 1 2007.
- [5] V. Maniezzo and A. Colomi, "The ant system applied to the quadratic assignment problem," *Ieee Transactions on Knowledge and Data Engineering*, vol. 11, pp. 769-778, Sep-Oct 1999.
- [6] M. Dorigo, *et al.*, "Ant system: Optimization by a colony of cooperating agents," *Ieee Transactions on Systems Man and Cybernetics Part B- Cybernetics*, vol. 26, pp. 29-41, Feb 1996.
- [7] P. R. McMullen, "An ant colony optimization approach to addressing a JIT sequencing problem with multiple objectives," *Artificial Intelligence in Engineering*, vol. 15, pp. 309-317, Jul 2001.
- [8] Colomi, *et al.*, "Ant system for job-shop scheduling," *Belgian Journal of Operations Research, Statistics and Computer Science*, vol. 34, pp. 39-53, 1994.
- [9] S. C. Chu, *et al.*, "Ant colony system with communication strategies," *Information Sciences*, vol. 167, pp. 63-76, Dec 2 2004.
- [10] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 53- 66, 1997.
- [11] A. Wade and S. Shalhi, "An ant system algorithm for the mixed vehicle routing problem with backhauls," in *Metaheuristics: Computer Decision-Making*, Norwell, MA: Kluwer, pp. 699- 719, 2004.
- [12] J. Han, *et al.*, "Mining top-k frequent closed patterns without minimum support," in *Proceedings of the 2002 IEEE international conference on data mining*, pp. 211- 218, 2002.
- [13] T. P. Hong, *et al.*, "An ACS-based framework for fuzzy data mining," *Expert Systems with Applications*, vol. 36, pp. 11844-11852, Nov 2009.
- [14] R. S. Parpinelli, *et al.*, "An Ant Colony Based System for Data Mining: Application to Medical Data," *The Genetic and Evolutionary Computation Conference*, pp. 791- 798, 2001.
- [15] T. P. Hong, *et al.*, "Extracting membership functions in fuzzy data mining by Ant Colony Systems," *Proceedings of 2008 International Conference on Machine Learning and Cybernetics, Vols 1-7*, pp. 3979-3984, 2008.
- [16] C. H. Chen, *et al.*, "Cluster-based evaluation in fuzzy-genetic data mining," *IEEE Transactions on Fuzzy Systems*, vol. 16, pp. 249-262, Feb 2008.
- [17] "Microsoft Corporation. Example Database FoodMart of Microsoft Analysis Services."

# Enhancing K-Means Algorithm with Semi-Unsupervised Centroid Selection Method

R. Shanmugasundaram and Dr. S. Sukumaran

**Abstract**— The k-means algorithm is one of the frequently used clustering methods in data mining, due to its performance in clustering massive data sets. The final clustering result of the kmeans clustering algorithm is based on the correctness of the initial centroids, which are selected randomly. The original k-means algorithm converges to local minimum, not the global optimum. The k-means clustering performance can be enhanced if the initial cluster centers are found. To find the initial cluster centers a series of procedure is performed. Data in a cell is partitioned using a cutting plane that divides cell in two smaller cells. The plane is perpendicular to the data axis with very high variance and is intended to minimize the sum squared errors of the two cells as much as possible, while at the same time keep the two cells far apart as possible. Cells are partitioned one at a time until the number of cells equals to the predefined number of clusters, K. The centers of the K cells become the initial cluster centers for K-means. In this paper, an efficient method for computing initial centroids is proposed. A Semi Unsupervised Centroid Selection Method is used to compute the initial centroids. Gene dataset is used to experiment the proposed approach of data clustering using initial centroids. The experimental results illustrate that the proposed method is very much apt for the gene clustering applications.

**Index Terms**— Clustering algorithm, K-means algorithm, Data partitioning, initial cluster centers, semi-supervised gene selection.

## I. INTRODUCTION

CLUSTERING, or unsupervised classification, will be considered as a mixture of problem where the aim is to partition a set of data object into a predefined number of clusters [13]. Number of clusters might be established by means of the cluster validity criterion or described by user. Clustering problems are broadly used in many applications, such as customer segmentation, classification, and trend analysis. For example, consider that customers purchased a retail database records containing items. A clustering method could group the customers in such a way that customers with similar buying patterns are in the same cluster. Several real-world applications deal with high dimensional data. It is always a challenge for clustering algorithms because of the manual processing is practically not possible. A high quality computer-based clustering removes the unimportant features and replaces the original set by a smaller representative set of data objects.

K-means is a well known prototype-based [14], partitioning clustering technique that attempts to find a user-specified

number of clusters (K), which are represented by their centroids.

The K-means algorithm is as follows:

1. Select initial centers of the K clusters. Repeat the steps 2 through 3 until the cluster membership stabilizes.
2. Generate a new partition by assigning each the data to its closest cluster centers.
3. Compute new cluster centers as centroids of the clusters.

Though K-means is simple and can be used for a wide variety of data types, it is quite sensitive to initial positions of cluster centers. The final cluster centroids may not be optimal ones as the algorithm can converge to local optimal solutions. An empty cluster can be attained if no points are allocated to the cluster during the assignment step. Therefore, it is important for K-means to have good initial cluster centers [15, 16]. In this paper a Semi-Unsupervised Selection Method (SCSM) is presented. The organization of this paper is as follows. In the next section, the literature survey is presented. In Section III, efficient semi-supervised centroid selection algorithm is presented. The experimental results and are presented in Section IV. Section V concludes the paper.

## II. LITERATURE SURVEY

Clustering statistical data has been studied from early time and lots of advanced models as well as algorithms have been proposed. This section of the paper provides a view on the related research work in the field of clustering that may assist the researchers.

Bradley and Fayyad together in [2] put forth a technique for refining initial points for clustering algorithms, in particular k-means clustering algorithm. They presented a fast and efficient algorithm for refining an initial starting point for a general class of clustering algorithms. The iterative techniques that are more sensitive to initial starting conditions were used in most of the clustering algorithms like K-means, and EM normally converges to one local minima. They implemented this iterative technique for refining the initial condition which allows the algorithm to converge to a better local minimum value. The refined initial point is used to evaluate the performance of K-means algorithm in clustering the given data set. The results illustrated that the refinement run time is significantly lower than the time required to cluster the full database. In addition, the method is scalable and can be coupled with a scalable clustering algorithm to concentrate on the large-scale clustering problems especially in case of data mining.

Yang et al. in [3] proposed an efficient data clustering algorithm. It is well known that K-means (KM) algorithm is one of the most popular clustering techniques because it is

---

R. Shanmugasundaram, Associate Professor, Department of Computer Science, Erode Arts & Science College, Erode, India.

Dr. S. Sukumaran, Associate Professor, Department of Computer Science, Erode Arts and Science College, Erode, India.

unproblematic to implement and works rapid in most situations. But the sensitivity of KM algorithm to initialization makes it easily trapped in local optima. K-Harmonic Means (KHM) clustering resolves the problem of initialization faced by KM algorithm. Even then KHM also easily runs into local optima. PSO algorithm is a global optimization technique. A hybrid data clustering algorithm based on the PSO and KHM (PSOKHM) was proposed by Yang et al. in [3]. This hybrid data clustering algorithm utilizes the advantages of both the algorithms. Therefore the PSOKHM algorithm not only helps the KHM clustering run off from local optima but also conquer the inadequacy of the slow convergence speed of the PSO algorithm. They conducted experiments to compare the hybrid data clustering algorithm with that of PSO and KHM clustering on seven different data sets. The results of the experiments show that PSOKHM was simply superior to the other two clustering algorithms.

Huang in [4] put forth a technique that enhances the implementation of K-Means algorithm to various data sets. Generally, the efficiency of K-Means algorithm in clustering the data sets is high. The restriction for implementing K-Means algorithm to cluster real world data which contains categorical value is because of the fact that it was mostly employed to numerical values. They presented two algorithms which extend the k-means algorithm to categorical domains and domains with mixed numeric and categorical values. The k-modes algorithm uses a trouble-free matching dissimilarity measure to deal with categorical objects, replaces the means of clusters with modes, and uses a frequency-based method to modernize modes in the clustering process to decrease the clustering cost function. The k-prototypes algorithm, from the definition of a combined dissimilarity measure, further integrates the k-means and k-modes algorithms to allow for clustering objects described by mixed numeric and categorical attributes. The experiments were conducted on well known soybean disease and credit approval data sets to demonstrate the clustering performance of the two algorithms.

Kluger [5] first proposed spectral biclustering for processing gene expression data. But Kluger's focus is mainly on unsupervised clustering, not on gene selection.

There are some present works related to the finding initialization centroids.

1. Compute mean ( $\mu_j$ ) and standard deviation ( $\sigma_j$ ) for every  $j$ th attribute values.
2. Compute percentile  $Z_1, Z_2, \dots, Z_k$  corresponding to area under the normal curve from  $-\infty$  to  $(2s-1)/2k$ ,  $s=1, 2, \dots, k$  (clusters).
3. Compute attribute values  $x_s = z_s \sigma_j + \mu_j$  corresponding to these percentiles using mean and standard deviation of the attribute.
4. Perform the K-means to cluster data based on  $j$ th attribute values using  $x_s$  as initial centers and assign cluster labels to every data.
5. Repeat the steps of 3-4 for all attributes (1).
6. For every data item  $t$  create the string of the class labels  $P_t = (P_1, P_2, \dots, P_l)$  where  $P_j$  is the class label of  $t$  when using the  $j$ th attribute values for step 4 clustering.

7. Merge the data items which have the same pattern string  $P_t$  yielding  $K'$  clusters. The centroids of the  $K'$  clusters are computed. If  $K' > K$ , apply Merge- DBMSDC (Density based Multi Scale Data Condensation) algorithm [6] to merge these  $K'$  clusters into  $K$  clusters.

8. Find the centroids of  $K$  clusters and use the centroid as initial centers for clustering the original dataset using  $K$  Means.

Although the mentioned initialization algorithms can help finding good initial centers for some extent, they are quite complex and some use the K-Means algorithm as part of their algorithms, which still need to use the random method for cluster center initialization. The proposed approach for finding initial cluster centroid is presented in the following section.

### III. METHODOLOGY

#### 3.1. Initial Cluster Centers Deriving from Data Partitioning

The algorithm follows a novel approach that performs data partitioning along the data axis with the highest variance. The approach has been used successfully for color quantization [7]. The data partitioning tries to divide data space into small cells or clusters where intercluster distances are large as possible and intracluster distances are small as possible.

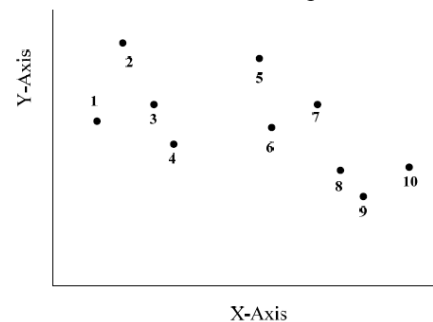


Fig. 1 Diagram of ten data points in 2D, sorted by its X value, with an ordering number for each data point

For instance, consider Fig. 1. Suppose ten data points in 2D data space are given.

The goal is to partition the ten data points in Fig. 1 into two disjoint cells where sum of the total clustering errors of the two cells is minimal, see Fig. 2. Suppose a cutting plane perpendicular to X-axis will be used to partition the data. Let  $C_1$  and  $C_2$  be the first cell and the second cell respectively and  $\bar{c}_1$  and  $\bar{c}_2$  be the cell centroids of the first cell and the second cell, respectively. The total clustering error of the first cell is thus computed by:

$$\sum_{c_i \in C_1} d(c_i, \bar{c}_1) \quad (1)$$

and the total clustering error of the second cell is thus computed by:

$$\sum_{c_i \in C_2} d(c_i, \bar{c}_2) \quad (2)$$

where  $c_i$  is the  $i^{\text{th}}$  data in a cell. As a result, the sum of total clustering errors of both cells are minimal (as shown in Fig. 2.)

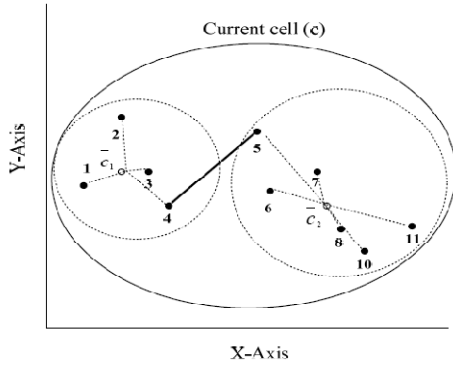


Fig. 2 Diagram of partitioning a cell of ten data points into two smaller cells, a solid line represents the intercluster distance and dash lines represent the intracluster distance

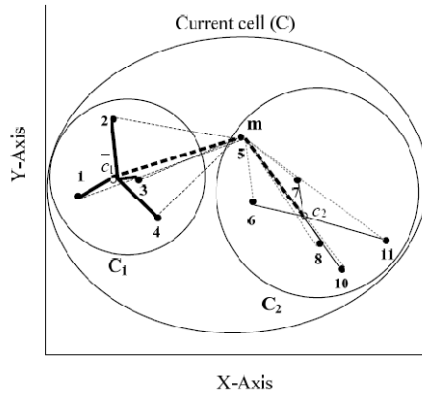


Fig. 3 Illustration of partitioning the ten data points into two smaller cells using m as a partitioning point. A solid line in the square represents the distance between the cell centroid and a data in cell, a dash line represents the distance between m and data in each cell and a solid dash line represents the distance between m and the data centroid in each cell

The partition could be done using a cutting plane that passes through m. Thus

$$d(c_i, \bar{c}_1) \leq d(c_i, c_m) + d(\bar{c}_1, c_m) \text{ and} \\ d(c_i, \bar{c}_2) \leq d(c_i, c_m) + d(\bar{c}_2, c_m) \quad (3)$$

(as shown in Fig. 3). Thus

$$\sum_{c_i \in C_1} d(c_i, \bar{c}_1) \leq \sum_{c_i \in C_1} d(c_i, c_m) + d(\bar{c}_1, c_m) \cdot |C_1| \\ \sum_{c_i \in C_2} d(c_i, \bar{c}_2) \leq \sum_{c_i \in C_2} d(c_i, c_m) + d(\bar{c}_2, c_m) \cdot |C_2| \quad (4)$$

m is called as the partitioning data point where  $|C_1|$  and  $|C_2|$  are the numbers of data points in cluster  $C_1$  and  $C_2$  respectively. The total clustering error of the first cell can be minimized by reducing the total discrepancies between all data in first cell to m, which is computed by:

$$\sum_{c_i \in C_1} d(c_i, c_m) \quad (5)$$

The same argument is also true for the second cell. The total clustering error of second cell can be minimized by reducing the total discrepancies between all data in second cell to m, which is computed by:

$$\sum_{c_i \in C_2} d(c_i, c_m) \quad (6)$$

where  $d(c_i, c_m)$  is the distance between m and each data in each cell. Therefore the problem to minimize the sum of total clustering errors of both cells can be transformed into the problem to minimize the sum of total clustering error of all data in the two cells to m.

The relationship between the total clustering error and the clustering point may be illustrated in Fig. 4, where the horizontal-axis represents the partitioning point that runs from 1 to n where n is the total number of data points and the vertical-axis represents the total clustering error. When  $m=0$ , the total clustering error of second cell equals to the total clustering error of all data points while the total clustering error of first cell is zero. On the other hand, when  $m=n$ , the total clustering error of the first cell equals to the total clustering error of all data points, while the total clustering error of the second cell is zero.

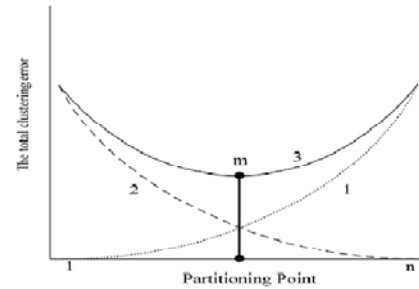


Fig. 4 Graphs depict the total clustering error, lines 1 and 2 represent the total clustering error of the first cell and second cell, respectively, Line 3 represents a summation of the total clustering errors of the first and the second cells

A parabola curve shown in Fig. 4 represents a summation of the total clustering error of the first cell and the second cell, represented by the dash line 2. Note that the lowest point of the parabola curve is the optimal clustering point (m). At this point, the summation of total clustering error of the first cell and the second cell are minimum.

Since time complexity of locating the optimal point m is  $O(n^2)$ , the distances between adjacent data is used along the X-axis to find the approximated point of n but with time of  $O(n)$ .

Let  $D_j = d(c_j, c_{j+1})^2$  be the squared Euclidean distance of adjacent data points along the X-axis.

If i is in the first cell then  $d(c_m, c_i) \leq \sum_{j=i}^m D_j$ . On the one hand, if i is in the second cell then  $d(c_m, c_i) \leq \sum_{j=m}^n D_j$  (as shown in Fig. 5).

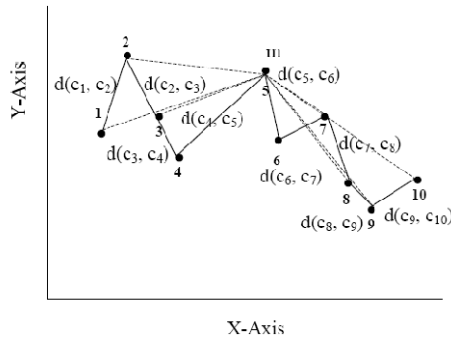


Fig. 5 Illustration of ten data points, a solid line represents the distance between adjacent data along the X-axis and a dash line represents the distance between m and any data point

The task of approximating the optimal point (m) in 2D is thus replaced by finding m in one-dimensional line as shown in Fig. 6.

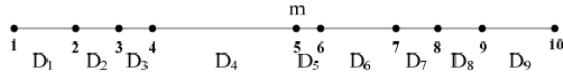


Fig. 6 Illustration of the ten data points on a one-dimensional line and the relevant  $D_j$

The point (m) is therefore a centroid on the one dimensional line (as shown in Fig. 6), which yields

$$\sum_{i=1}^{m-1} d(c_m, c_i) \approx \sum_{i=m}^n d(c_m, c_i) \quad (7)$$

Let  $dsum_i = \sum_{j=1}^i D_j$  and a *centroidDist* can be computed

$$centroidDist = \frac{\sum_{i=1}^n dsum_i}{n} \quad (8)$$

It is probable to choose either the X-axis or Y-axis as the principal axis for data partitioning. However, data axis with the highest variance will be chosen as the principal axis for data partitioning. The reason is to make the inter distance between the centers of the two cells as large as possible while the sum of total clustering errors of the two cells are reduced from that of the original cell. To partition the given data into k cells, it is started with a cell containing all given data and partition the cell into two cells. Later on the next cell is selected to be partitioned that yields the largest reduction of total clustering errors (or Delta clustering error). This can be described as Total clustering error of the original cell – the sum of Total clustering errors of the two sub cells of the original cell. This is done so that every time a partition on a cell is performed, the partition will help reduce the sum of total clustering errors for all cells, as much as possible.

The partitioning algorithm can be used now to partition a given set of data into k cells. The centers of the cells can then be used as good initial cluster centers for the K-means algorithm. Following are the steps of the initial centroid predicting algorithm.

1. Let cell c contain the entire data set.
2. Sort all data in the cell c in ascending order on each attribute value and links data by a linked list for each attribute.

3. Compute variance of each attribute of cell c. Choose an attribute axis with the highest variance as the principal axis for partitioning.

4. Compute squared Euclidean distances between adjacent data along the data axis with the highest variance  $D_j = d(c_j, c_{j+1})^2$  and compute the  $dsum_i = \sum_{j=1}^i D_j$

5. Compute centroid distance of cell c:

$$centroidDist = \frac{\sum_{i=1}^n dsum_i}{n}$$

Where  $dsum_i$  is the summation of distances between the adjacent data.

6. Divide cell c into two smaller cells. The partition boundary is the plane perpendicular to the principal axis and passes through a point m whose  $dsum_i$  approximately equals to *centroidDist*. The sorted linked lists of cell c are scanned and divided into two for the two smaller cells accordingly

7. Calculate Delta clustering error for c as the total clustering error before partition minus total clustering error of its two sub cells and insert the cell into an empty Max heap with Delta clustering error as a key.

8. Delete a max cell from Max heap and assign it as a current cell.

9. For each of the two sub cells of c, which is not empty, perform step 3 - 7 on the sub cell.

10. Repeat steps 8 - 9. Until the number of cells (Size of heap) reaches K.

11. Use centroids of cells in max heap as the initial cluster centers for K-means clustering

The above presented algorithms for finding the initialization centroids do not provide a better result. Thus an efficient method is proposed for obtaining the initial cluster centroids. The proposed approach is well suited to cluster the gene dataset. So the proposed method is explained on the basis of genes.

### 3.2. Proposed Methodology

The proposed method is Semi-Unsupervised Centroid Selection method. The proposed algorithm finds the initial cluster centroids for the microarray gene dataset. The steps involved in this procedure are as follows.

Spectral biclustering [10-12] can be carried out in the following three steps: data normalization, Bistochasticization, and seeded region growing clustering. The raw data in many cancer gene-expression datasets can be arranged in one matrix. In this matrix, denoted by, the rows and columns represent the genes and the different conditions (e.g., different patients), respectively. Then the data normalization is performed as follows. Take the logarithm of the expression data. Carry out five to ten cycles of subtracting either the mean or median of the rows (genes) and columns (conditions) and then perform five to ten cycles of row-column normalization. Since gene expression microarray experiments can generate data sets with multiple missing values, the k-nearest neighbor (KNN) algorithm is used to fill those missing values.

Define  $\bar{A}_i = (1/m) \sum_{j=1}^m A_{ij}$  to be the average of ith row,  $\bar{A}_i = (1/n) \sum_{j=1}^n A_{ij}$  to be the average of th column, and



$\bar{A}_{..} = (1/mn) \sum_{j=1}^n \sum_{i=1}^m A_{ij}$  to be the average of the whole matrix, where m is the number of genes and n the number of conditions.

Bistochastization may be done as follows. First, a matrix of interactions is defined  $K = (K_{ij})$  by  $K_{ij} = A_{ij} - \bar{A}_{i.} - \bar{A}_{.j} + \bar{A}_{..}$ . Then the singular value decomposition (SVD) of the matrix K is computed as given by  $K = U\Lambda V^T$ , where  $\Lambda$  is a diagonal matrix of the same dimension as K and with nonnegative diagonal elements in decreasing order, U and V are  $m \times m$  and  $n \times n$  orthonormal column matrices. The th column of the matrix V is denoted by  $\vec{v}_1$  and  $\vec{v}_2$ . Therefore, a scatter plot of experimental conditions of the two best class partitioning eigenvectors  $\vec{v}_2$  and  $\vec{v}_2$  is obtained. The  $\vec{v}_1$  and  $\vec{v}_2$  are often chosen as the eigenvectors corresponding to the largest and the second largest eigenvalues, respectively. The main reason is that they can capture most of the variance in the data and provide the optimal partition of different experimental conditions. In general, an s-dimensional scatter plot can be obtained by using eigenvectors  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_s$  (with largest eigenvalues).

Define  $P = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_s]^T$  which has a dimension of  $n \times s$ . The rows of matrix P stand for different conditions, which will be clustered using SRG. Seeded region growing clustering is carried out as follows. It begins with some seeds (initial state of the clusters). At each step of the algorithm, it is considered all as-yet unallocated samples, which border with at least one of the regions. Among them one sample, which has the minimum difference from its adjoining cluster, is allocated to its most similar adjoining cluster. With the result of clustering, the distinct types of cancer data can be predicted with very high accuracy. In the next section, such clustering result is used to select the best gene combinations or explained as the best initial centroids.

### 3.2.1. Semi-Unsupervised Centroid Selection (SCSM)

The proposed semi-supervised centroid selection method includes two steps: gene ranking and gene combination selection.

As stated above, the best class partitioning eigenvectors is obtained. Now these eigenvectors  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_s$  are used to rank and preselect genes.

The proposed semi-supervised centroid selection method is based on the following two assumptions.

- The genes which are most relevant to the cancer should capture most variance in the data.
- Since  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_s$  may reveal the most variance in the data, the genes "similar" to  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_s$  should be relevant to the cancer

The gene ranking and preselecting process can be summarized as follows. After defining the ith gene profile  $\vec{g}_i = (a_{i1}, a_{i2}, \dots, a_{in})$ , cosine measure is used to

compute the correlation (similarity) between each gene profile (e.g.,) and the eigenvectors  $(e.g.) \vec{v}_j, j = 1, 2, \dots, s$  as

$$R_{i,j} = \frac{(\vec{g}_i)^T \vec{v}_j}{\|\vec{g}_i\|_2 \cdot \|\vec{v}_j\|_2}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, s \quad (9)$$

Where  $\|\cdot\|_2$  means vector 2—norms. Seen from (10), a large absolute of  $R_{i,j}$  indicates a strong correlation (similarity) between ith gene and jth eigenvector. Therefore, genes can be ranked as the absolute correlation values  $|R_{i,j}|$  for each eigenvector. For the eigenvector the top genes can be preselected, denoted by  $G_j$ , according to the corresponding  $|R_{i,j}|$  value for  $j = 1, 2, \dots, s$ . The value l can be empirically determined. Thus, for each eigenvector of  $\vec{v}_1, \dots, \vec{v}_s$  a set of genes with largest values of the Cosine Measure is obtained which are taken as the initial cluster centroids in the proposed clustering technique.

## IV. EXPERIMENTAL RESULTS

The proposed SCSM method is experimented using two microarray data sets: the lymphoma data set and the liver cancer data set.

TABLE I  
GENE IDS (CLIDS) AND GENE NAMES IN THE TWO MICROARRAY DATA SETS

Data set	Gene ID/ CLID	Gene Name	Gene Rank	
			G1	G2
Lymphoma	GENE 1622X	*CD63 antigen (melanoma 1 antigen); Clone=769861	3	/
	GENE 2328X	*FGR tyrosine Kinase; Clone=728609	/	3
	GENE 3343X	*mosaic protein LR11=hybrid; Receptor gp250 precursor; Clone=1352833	/	4
Liver Cancer	IMAGE: 301122	116682 ECM1 extracellular matrix protein 1 Hs.81071 N79484	7	/

The lymphoma microarray data has three subtypes of cancer, i.e., CLL, FL, and DLCL. The dataset is obtained from [8]. When applying the proposed method to this data set, the clustering result with two best partition eigenvectors is obtained. Seen from cluster results the three classes are correctly divided. Then two sets of l=20 genes are selected

TABLE II  
COMPARISON OF GENERALIZATION ABILITY

Data set	Method	Number of genes selected	Test Rate (%)	( <i>p1,p2</i> )
Lymphoma	k-means	4026	100±0	(0, 0.9937)
	Existing Method	81	100±0	(0, 0.9937)
	SCSM	2±0	99.92±0.37	(NA, NA)
Liver Cancer	k-means	1648	98.10±0.11	(0, 0.9973)
	Existing Method	23	98.10±0.11	(0, 0.9973)
	SCSM	1±0	98.70±0.08	(NA, NA)

according to  $|R_{i,1}|$  and  $|R_{i,2}|$  respectively. (Here set have to be two.) From the two sets of 20 genes each, the two-gene combinations is chosen that can best divide the lymphoma data. Two pairs of genes have been found: 1) Gene 1622X and Gene 2328X, and 2) Gene 1622X and Gene 3343X, which perfectly divide the lymphoma data. Since the results are similar to each other, only the result of one group is shown. Gene ID and gene names of the selecting genes in the lymphoma data set are shown in Table I, where the group and the rank of genes are also shown.

The method is applied to the liver cancer data with two classes, i.e., nontumor liver and HCC. The lung cancer data is obtained from [9]. The clustering result with the two best partition eigenvectors is obtained. From the results it can be seen that there are three samples misclassified and the clustering accuracy is 98.1%. Actually, it can set so that the scatter plot is on a single axis. Then top 20 genes are selected with the largest. From the top 20 genes, it is found one gene that can divide the liver cancer data well with accuracy of 98.7%. The result and gene name of selecting gene in liver cancer data set are shown in Table I.

#### 4.1. Comparison with results

The paired t-test method is used to show the statistical difference between our results and other published results. In general, given two paired sets and of measured values, the paired t-test can be employed to compute a  $-value$  between and determines whether they differ from each other in a statistically significant way under the assumptions that the paired differences are independent and identically normally distributed. The  $-value$  is defined as follows:

$$p = (\bar{X} - \bar{Y}) \sqrt{\frac{n(n-1)}{\sum_{i=1}^n (\hat{X}_i - \hat{Y}_i)^2}}$$

Where,  $\hat{X}_i = X_i - \bar{X}$ ,  $\hat{Y}_i = Y_i - \bar{Y}$  and  $\bar{X}, \bar{Y}$  are the mean values for and , respectively. Hence, all  $p \in [0,1]$ , with a high  $-value$  indicating statistically insignificant differences and a low  $-value$  indicating statistically significant differences between  $X_i$  and  $Y_i$ .

The order of cancer subtypes are shuffled and carried out the experiments 20 times for each data set. Each time the

same gene selection result is obtained for each data set, but slightly different classification accuracies. The p-values for both numbers of genes and classification accuracies is calculated for both data sets in Table II, which showed that the differences between the numbers of genes used in our method and other methods are statistically significant, whereas the differences between the classification accuracies between the proposed method and other methods are not statistically significant.

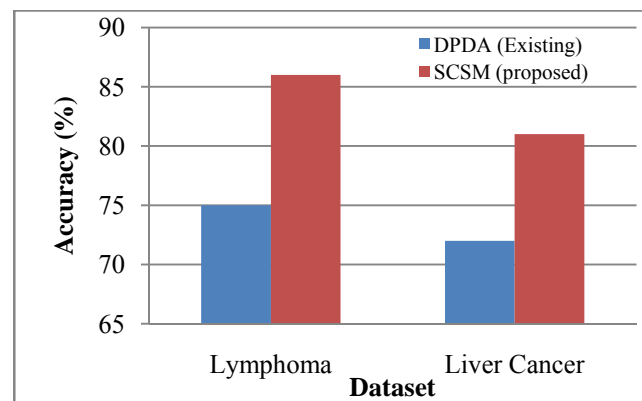


Figure 7: Comparison of classification accuracy among the proposed and existing technique for two different datasets.

The Figure 7 shows that the DPDA-K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance method produces result with less percentage of accuracy than the proposed clustering with SCSM. The classification accuracy of the proposed method is very high than all the existing method. The result also shows that the proposed method is suitable only for the gene clustering and when the proposed method used to cluster the other data it produces a less percentage of accuracy.

The figure 8 shows the comparison of clustering time among the DPDA-K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance method and the proposed clustering with SCSM.

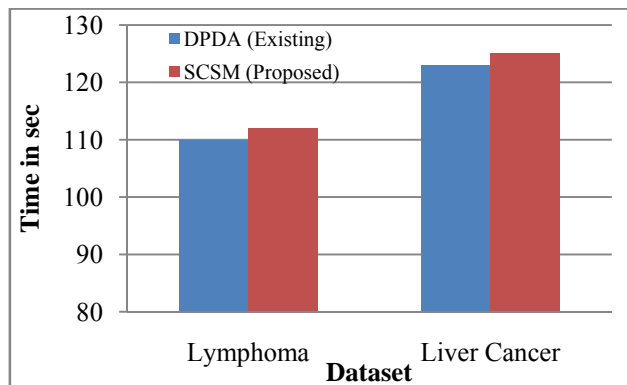


Figure 8: Comparison of classification time among the proposed and existing technique for two different datasets.

From the graph it can be easily said that the proposed method takes slightly more time to cluster the gene data than the existing method. Even the clustering time taken is more, clustering accuracy is very high. Thus the proposed system can be used for the gene clustering.

## V. CONCLUSION

The most commonly used efficient clustering technique is k-means clustering. Initial starting points those computed randomly by K-means often make the clustering results reaching the local optima. So to overcome this disadvantage a new technique is proposed. Semi-Unsupervised Centroid Selection method is used with the present clustering approach in the proposed system to compute the initial centroids for the k-means algorithm. The experiments for this proposed approach is conducted on the microarray gene database. The data sets used are lymphoma and the liver cancer data set. The accuracy of the proposed approach is compared with the existing technique called the DPDA. The results are obtained and tabulated. It is clearly observed from the results that, the proposed approach shows significant performance. In the lymphoma data set, the accuracy of the proposed approach is about 87%. The accuracy of the DPDA approach is very less (i.e.) 75%. Similarly for the liver cancer data set, the accuracy of the proposed approach is about 81% which is also higher than the existing approach. Moreover, time taken for classification of the proposed approach is more or less similar to the DPDA approach. The time taken for classification by the proposed approach in lymphoma and liver cancer data sets are 115 and 130 seconds respectively which is almost similar to the existing approach. Thus the proposed approach provides the best classification accuracy within a short time interval.

## REFERENCES

[1] Guangsheng Feng, Huiqiang Wang, Qian Zhao, and Ying Liang, "A Novel Clustering Algorithm for Prefix-Coded Data Stream Based upon Median-Tree," IEEE, International Conference on Internet Computing in Science and Engineering, ICICSE '08, pp. 79-84, 2008.  
[2] P. S. Bradley, and U. M. Fayyad, "Refining Initial Points for K-Means Clustering," ACM, Proceedings of the 15<sup>th</sup> International Conference on Machine Learning, pp. 91-99, 1998.

[3] F. Yang, T. Sun, and C. Zhang, "An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization," An International Journal on Expert Systems with Applications, vol. 36, no. 6, pp. 9847-9852, 2009.  
[4] Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values," Journal on Data Mining and Knowledge Discovery, Springer, vol. 2, no. 3, pp. 283-304, 1998.  
[5] Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, "Spectral biclustering of microarray cancer data: co-clustering genes and conditions," Genome Res., vol. 13, pp. 703-716, 2003.  
[6] P. Fränti and J. Kivijärvi, "Randomised Local Search Algorithm for the Clustering Problem". Pattern Analysis and Applications, Volume 3, Issue 4, pages 358 – 369, 2000.  
[7] M. Halkidi, Y. Batistakis and M. Vazirgiannis, "Cluster Validity Methods: part I". In Proceedings of the ACM SIGMOD International Conference on Management of Data, Volume 31, Issue 2, pages 40 – 45, June 2002.  
[8] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.  
[9] Hong, Z.Q. and Yang, J.Y. "Optimal Discriminant Plane for a Small Number of Samples and Design Method of Classifier on the Plane", Pattern Recognition, Vol. 24, No. 4, pp. 317-324, 1991.  
[10] Manjunath Aradhya, Francesco Masulli, and Stefano Rovetta "Biclustering of Microarray Data based on Modular Singular Value Decomposition", Proceedings of CIBB 2009.  
[11] LIU Wei, "A Parallel Algorithm for Gene Expressing Data Biclustering", journal of computers, vol. 3, no. 10, october 2008  
[12] Kenneth Bryan, P'adraig Cunningham and Nadia Bolshakova, "Biclustering of Expression Data Using Simulated Annealing", This research was sponsored by Science oundation Ireland under grant number SFI-02/IN1/1111.  
[13] A. K. Jain, M. N. Murty and P. J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, September 1999  
[14] Shai Ben-David, David Pal, and Hans Ulrich Simon, "Stability of k-Means Clustering".  
[15] Madhu Yedla, Srinivasa Rao Pathakota and T. M. Srinivasa, "Enhancing K-means Clustering Algorithm with Improved Initial Center", Vol. 1, 121-125, 2010.  
[16] A. M. Fahim, A. M. Salem, F. A. Torkey and M. A. Ramadan, "An Efficient enhanced k-means clustering algorithm," journal of Zhejiang University, 10(7): 16261633, 2006.

# A Survey on Static Power Optimization in VLSI

A. Janaki Rani and Dr. S. Malarkkan

**Abstract**---Power has become one of the primary constraints for both the high performance and portable system design. The growing market of battery powered electronic systems like cellular phones, personal digital assistants demands the design of microelectronic circuits with low power consumption. Power dissipation in these systems may be divided into two major components namely static and dynamic power dissipation. The static power is the standby power that is wasted even if the device is not performing any function. As technology scales down the static power dissipation is dominant in VLSI circuits which are mainly due to leakage current in transistors. Hence a focus is necessary on the leakage currents. These leakage currents are mainly due to sub-threshold leakage and gate oxide leakage. The sub-threshold leakage is dominant which can be minimized by reducing the supply voltage, reducing the transistor size, decreasing the temperature and increasing the threshold voltage. In this paper a survey is presented on static power optimization in VLSI. It presents the possible solutions to reduce the leakage power in various digital logic circuits like CMOS, I<sup>2</sup>C etc.

**Index Terms**—Leakage, Low-Power, Power Gating, Semicustom, Input Vector Control, Body Bias Control, Sleep Transistor Sizing, Sleepy Stack, Zigzag Power Gating (ZPG)

## I. INTRODUCTION

IN the past, the major concerns of the VLSI designer were Area, performance, cost and reliability; power considerations were mostly of only secondary importance. In recent years, however, this has begun to change and, increasingly, power is being given comparable weight to area and speed. Several factors have contributed to this trend. Portable computing and communication devices demand high-speed computation and complex functionality with low power consumption. Heat production in high-end computer products limits the feasible packing and performance of VLSI circuits and increases the packaging and cooling costs. Circuit and device reliability deteriorate with increased heat dissipation, and thus the die temperature. Heat pumped into the rooms, the electricity consumed and the office noise diminishes with low power LSI chipset. Leakage-power problems are a serious issue in portable electronic systems that operate mostly in standby mode. Lowering power-supply voltage in the system is one of the most effective schemes to reduce the power dissipation. As the VLSI technology and threshold/supply voltage continue scaling down, leakage power has become more and more significant in the power dissipation of today's CMOS circuits. For example, it is projected that subthreshold leakage power can contribute as much as 42% of the total power in the 90nm process generation [1].

The power dissipation can be minimized only if the source of power dissipation is analyzed. Power dissipation in digital CMOS circuits is caused due to sources as follows. (a) The leakage current, which is primarily found by the fabrication technology, consists of four components namely sub-threshold leakage current ( $I_{sub}$ ), gate direct tunneling current ( $I_g$ ), gate-induced drain leakage current ( $I_{GIDL}$ ) and reverse-biased junction leakage current ( $I_{rev}$ ), (b) the standby current which is the DC current drawn continuously from  $V_{dd}$  to ground, (c) the short-circuit (rush-through) current which is due to the DC path between the supply rails during output transitions, (d) the capacitance current which flows to charge and discharge capacitive loads during logic changes. The term static power dissipation describes the sum of leakage and standby dissipations. The static power dissipation is dominated by the leakage components and is given by

$$P_{static} = I_{leak} * V_{dd} \text{ ----- (1)}$$

The sub-threshold leakage current and gate direct tunneling current are dominant in the sub-100nm CMOS circuits. The sub-threshold leakage current is given by

$$I_{subth} = I_0 \exp[(V_{gs} - V_t) / (n V_T)] [1 - \exp(-V_{ds}/V_T)] \text{ --- (2)}$$
$$\text{And } I_0 = \mu_{eff} C_{ox} (W/L) V_T^2 \text{ ----- (3)}$$

Where  $\mu_{eff}$  is the electron/hole mobility,  $C_{ox}$  is the gate capacitance per unit area,  $W$  and  $L$  are width and length of the channel respectively,  $V_t$  is the threshold voltage,  $n$  is the sub-threshold swing co-efficient,  $V_T$  is the thermal voltage,  $V_{gs}$  is the transistor gate to source voltage and  $V_{ds}$  is the drain to source voltage. The sources of static power dissipation is summarized above and the following section of this paper presents the literature survey on the different techniques for reducing leakage current which are power supply gating, dual threshold voltage, input vector control, body bias control, sleepy stack, forced stacking and use of MTCMOS, VTCMOS and guarding.

## II. LITERATURE SURVEY

Mutoh et al, [3] presented the concept of Multi Threshold CMOS (MTCMOS). In this technique, a high-threshold voltage transistor is inserted in series with the power supply and the existing design and ground as shown in Figure 1. During active mode of operation, the high threshold  $V_t$  transistors are turned on, thereby facilitating normal operation of the circuit as there exists a direct path from the output to

---

A. Janaki Rani, Research Scholar Sathyabama University, Chennai, India.  
Dr. S. Malarkkan, Principal, Manakulavinayagar Institute of Technology,  
Pondicherry, India.

ground and  $V_{dd}$ . During standby mode, these transistors are turned off creating a virtual power supply and ground rail and cutting off the circuit from supply. The high  $V_t$  transistors are called sleep transistors.

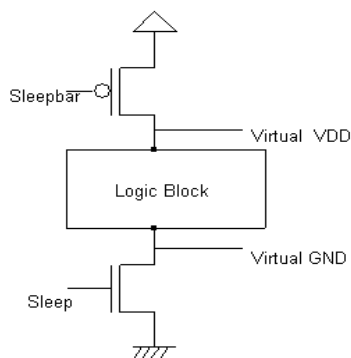


Fig.1 MTCMOS Circuit

Narendra et al, [4] showed that stacking of two off transistors significantly reduces sub-threshold leakage compared to a single off transistor. It is an effective way to reduce leakage power in active mode. Transistor stacking technique uses the dependence of  $I_{\text{sub}}$  on the source terminal voltage  $V_s$ . With the increase of  $V_s$  of the transistor, the sub-threshold leakage current reduces exponentially. If natural stacking of transistors does not exist in a circuit, then to utilize the stacking effect a single transistor of width  $W$  is replaced by two transistors each of width  $W/2$ . This is called forced stacking as shown in Figure 2.

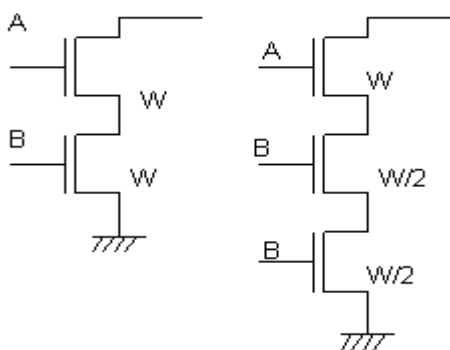


Fig. 2 Forced stacking circuit

J.C. Park et al, [5] described a sleepy stack technique which combines the sleep transistor approach during active mode and the stack approach during standby mode. In this technique, forced stacking is first implemented. Then to one of the stacked transistors a sleep transistor is inserted in parallel. Thus during active mode, the sleep transistors are on thereby reducing the effective resistance of the path. This leads to reduced propagation delay during active mode as compared to the forced stacking method. During standby mode, the sleep transistor is turned off and the stacked transistor suppresses the leakage power. Figure 3 shows the circuit of a sleepy stack inverter, where the S and S' are sleep control signals.

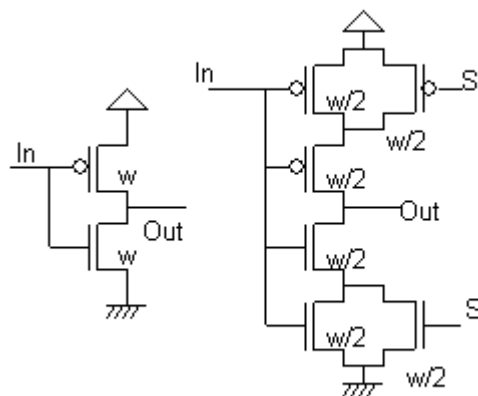


Fig.3 Sleepy stack inverter circuit

Preetham Lakshmikanthan et al, [6] presented a self-controlling leakage reduction technique called VCLEARIT for CMOS circuits. Signal probabilities determine the mode of operation (active or standby) of the gates making up complex circuits. Cancellation of leakage effects in both the pull up network (PUN) as well as the pull down network for CMOS gates results in leakage reduction. A combination of high  $V_t$  and standard  $V_t$  control transistors achieve voltage balancing in the pull up and pull down paths. Results show up to 61% reduction in leakage power of combinational circuits.

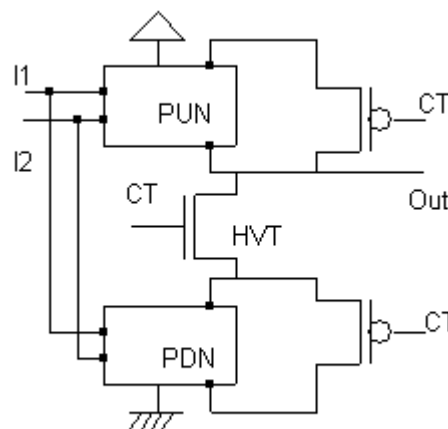


Fig. 4 VCLEARIT Circuit

Yu et al, [7] proposed a power optimization technique on reduced leakage power with thermal awareness using dual threshold voltage. Dual-  $V_{th}$  design is an effective leakage power reduction technique at behavioral synthesis level. It permits designers to replace modules on non-critical path with the high- $V_{th}$  implementation. Though, the existing constructive algorithms fail to find the optimal solution due to the complexity of the problem and do not consider the on-chip temperature variation. In his research, a two-stage thermal dependent leakage power minimization algorithm is proposed by using dual-  $V_{th}$  library during behavioral synthesis. In the

first stage, the timing impact on other modules caused by replacing certain modules with high  $V_{th}$  is quantitatively evaluated. Based on this analysis and the characteristics of the dual-  $V_{th}$  module library, a small set of candidate solutions is generated for the module replacement. Then in the second stage, the on-chip thermal information is obtained from thermal-aware floor planning and thermal analysis to select the final solution from the candidate set. Experimental results show an average of 17.8% saving in leakage energy consumption and a slightly shorter runtime compared to the best known work. In most cases, this algorithm can actually find the optimal solution obtained from a complete solution space exploration.

One of the most effective design techniques for reducing leakage is dual-  $V_{th}$  design [8], where performance-critical transistors are made of low-  $V_{th}$  to provide the required performance and high-  $V_{th}$  transistors are used everywhere else to reduce leakage. Dual-  $V_{th}$  assignment or allocation can be applied to all phases of the design flow. Although transistor level dual-  $V_{th}$  allocation is the most effective for leakage reduction, it is also the most challenging due to the complexity of dealing with the billions of transistors in modern ICs. Thus it has been proposed [9, 10] to allocate  $V_{th}$  at behavioral level, where the solution space is much smaller than that at the transistor level. At behavioral level, dual-  $V_{th}$  allocation can be converted to the module selection problem. The modules on noncritical path are selected to be replaced with high-  $V_{th}$  implementation. Due to factors such as module sharing in behavioral level, it is difficult to model the timing relationship of modules precisely. And consequently, the optimal module selection will be hard to obtain. Kawaguchi et al, [11] proposed a circuit called Super Cutoff CMOS (SCCMOS), which is an alternate to MTCMOS power gating. In this technique, the sleep transistors are under-driven (NMOS) or over-driven (PMOS) when in the standby mode. An NMOS transistor will be turned off with a slight negative gate voltage instead of zero voltage. This negative gate voltage decreases the sub-threshold leakage current exponentially. In this scheme the sleep transistor and the logic transistors are having the same standard  $V_t$ . Therefore the circuit operates fast in the active mode. During standby mode, since the transistor is turned off with a negative gate voltage, sub-threshold leakage current reduces exponentially.

Yuan et al, [12] presented an Input Vector Control approach for leakage current reduction. IVC takes advantage of transistor stack effect to apply the minimum leakage vector (MLV) to the primary inputs of the circuit during the standby mode. Though, IVC technique becomes less effective for circuits of large logic depth because the MLV at primary inputs has little impact on internal gates at high logic level. In his research, a technique is presented to overcome this limitation by directly controlling the inputs to the internal gates that are in their worst leakage states. Specifically, a gate replacement technique is proposed that replaces such gates by other library gates while maintaining the circuit's correct functionality at the active mode. This alteration of the circuit does not require changes of the design flow, but it opens the

door for further leakage reduction, when the MLV is not effective. The author then describes a divide and-conquer approach that combines the gate replacement and input vector control techniques. It incorporates an algorithm that finds the optimal MLV for tree circuits, a fast gate replacement heuristic, and a genetic algorithm that connects the tree circuits.

Behnam et al, [13] discussed on the leakage minimization of SRAM Cells in a Dual- $V_t$  and Dual- $T_{ox}$  Technology. Aggressive CMOS scaling results in the low threshold voltage and thin oxide thickness for transistors manufactured in deep submicron regime. As a result, reducing subthreshold and tunneling gate leakage currents has become one of the most important criteria in the design of VLSI circuits. His research puts forth a method based on dual- $V_t$  and dual- $T_{ox}$  assignment to reduce the total leakage power dissipation of SRAMs while maintaining their performance. The technique is based on the observation that read and writes delays of a memory cell in an SRAM block depend on the physical distance of the cell from the sense amplifier and the decoder. Thus, the idea is to implement different configurations of six-transistor SRAM cells corresponding to different threshold voltage and oxide thickness assignments for the transistors. Different to other techniques for low leakage SRAM design, the proposed technique incurs neither area nor delay overhead. In addition, it results in a minor change in the SRAM design flow. The leakage saving obtained by using this technique is a function of the values of the high threshold voltage and the oxide thickness, as well as the number of rows and columns in the cell array.

CMOS scaling beyond the 90nm technology node requires not only very low threshold voltages ( $V_t$ ) to retain the device switching speeds, but also ultra-thin gate oxides ( $T_{ox}$ ) to maintain the current drive and keep threshold voltage variations under control when dealing with short-channel effects [14]. Low threshold voltage results in an exponential increase in the subthreshold leakage current, whereas ultra-thin oxide causes an exponential increase in the tunneling gate leakage current. The leakage power dissipation is approximately proportional to the area of a circuit. Since in many processors caches occupy about 50% of the chip area [15], the leakage power of caches is one of the major sources of power consumption in high performance microprocessors.

While one of the ways in reducing the subthreshold leakage is to use higher threshold voltages in some parts of a design, to suppress tunneling gate leakage, high-k dielectrics or multiple gate oxides may be used. In [16, 17] a comparative study of using high-k dielectric and dual oxide thickness on the leakage power consumption has been presented and an algorithm for simultaneous high-k and high- $T_{ox}$  assignment has been proposed. Although some investigation has been done on Zirconium- and Hafnium-based high-k dielectrics [18], there are unresolved manufacturing process challenges in way of introducing high-k dielectric material under the gate (e.g., related to the compatibility of these materials with Silicon [19] and the need to switch to metal gates); hence, high-k dielectrics are not expected to be used before 45nm



technology node [18,20], leaving multiple gate oxide thicknesses as the one promising solution to reduce tunneling gate leakage current at the present time. Kyung Ki Kim et al, [21] proposed a novel design method to minimize the leakage power during standby mode using a novel adaptive supply voltage and body-bias voltage generating technique. Based on the temperature and process conditions, the optimal supply voltage is generated to reduce leakage power. The body bias voltage is automatically adjusted continuously by the control loop to adapt to the process voltage and temperature (PVT) variations. By tuning body-bias voltage using leakage monitoring circuit, circuits can be biased at the optimal point where sub-threshold leakage current and band-to-band-tunneling (BTBT) leakage current are balanced to accomplish the minimum leakage power.

Changbo et al, [23] puts forth a Distributed Sleep Transistor Network for power reduction. Sleep transistors are efficient to reduce dynamic and leakage power. The cluster-based design (Refer Fig. 4) was presented to reduce the sleep transistor area by clustering gates to minimize the simultaneous switching current per cluster and then inserting a sleep transistor per cluster. In the research, the author proposes a novel distributed sleep transistor network (DSTN), and show that DSTN is intrinsically better than the cluster based design in terms of the sleep transistor area and circuit performance. The author reveals properties of optimal DSTN designs, and then develops an efficient algorithm for gate level DSTN synthesis. The algorithm obtains DSTN designs with up to 70.7% sleep transistor area reduction when compared to cluster-based designs. Furthermore, the author presents custom layout designs to verify the area reduction by DSTN. In the cluster-based structure shown in Figure 4, a module is decomposed into several logic clusters, and each cluster is supported by one local sleep transistor. Figure 5 shows the distributed structure called Distributed Sleep Transistor Network (DSTN). In the DSTN structure, the cluster-based sleep transistor deployment is enhanced by connecting all the virtual ground lines (VGND) together, thus allowing the operating current from each cluster to flow through all the sleep transistors. In this way, the discharged current among the sleep transistors tends to be balanced.

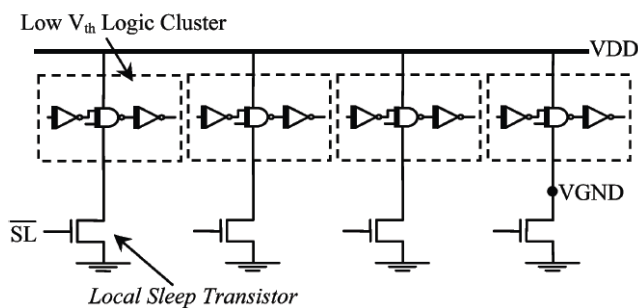


Fig. 5. Cluster based design

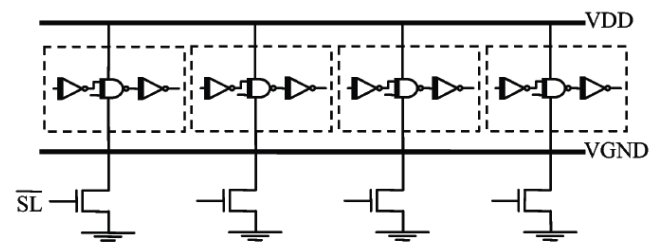


Fig. 6 Distributed Sleep Transistor Network (DSTN)

Narender Hanchate et al, [23] proposed a novel technique called LECTOR (Refer Fig. 6) for reducing leakage power in CMOS circuits. He introduced two leakage control transistors (LCT) a PMOS and NMOS within the logic gate. The gate terminal of each LCT is controlled by the source of the other. In this arrangement, one of the LCT's is always near its cut-off voltage for any input combination. This increases the resistance of the path from  $V_{dd}$  to ground leading to significant decrease in leakage currents. This technique works effectively in both active and idle states of the circuit, resulting in better leakage reduction. The experimental results indicate an average leakage reduction of 79.4% for MCNC "91 benchmark circuits. De-Shiuan et al, [24] discussed on the power reduction using sleep transistor sizing for leakage power minimization considering charge balancing. One of the efficient techniques to reduce leakage power is power gating. Previously, a DSTN was proposed to reduce the sleep transistor area for power gating by connecting all the virtual ground lines together to minimize the Maximum Instantaneous Current flowing through sleep transistors. In his research, a new methodology is proposed for determining the sizes of sleep transistors of the DSTN structure. The author presents novel algorithms and theorems for efficiently estimating a tight upper bound of the voltage drop and minimizing the sizes of sleep transistors. The author also presents mathematical proofs of the theorems and lemmas in detail. The experimental results show 23.36% sleep transistor area reduction when compared to the previous work on space reduction.

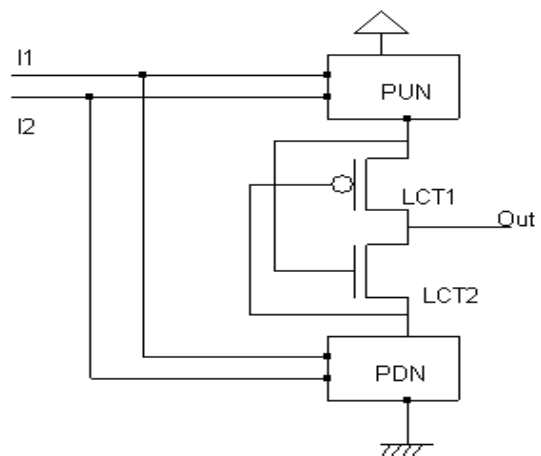


Fig. 7 LECTOR Circuit

Youngsoo et al, [25] presented a Semicustom Design of Zigzag Power-Gated Circuits in Standard Cell Elements. ZPG solved the long wake-up delay of standard power gating, but its requirement for both nMOS and pMOS current switches, in a zigzag pattern, requires complicated power networks, limiting application to custom designs. The Zigzag Power gating (ZPG) circuit is shown in Figure 7. The author proposed a design framework for cell-based semicustom design of ZPG circuits, using a new power network architecture that allows the unmodified conventional logic cells to be combined with custom circuitry such as ZPG flip-flops, input forcing circuits, and current switches. The design flow, from the register transfer level description to layout, is described and applied to a 32-b microprocessor design using a 1.2-V 65-nm triple-well bulk CMOS process. The use of a sleep vector in ZPG needs additional switching power when entering standby mode and returning to active mode. The switching power must be minimized so that it does not outweigh the leakage saved by employing ZPG scheme. The author formulates the selection of a sleep vector as a multiobjective optimization problem, minimizing both the transition energy and the total wire length of a design. The author solved the problem by employing multiobjective genetic-based algorithm. Experimental results of the author technique show an average saving of 39% in transition energy and 8% in total wire length for several benchmark circuits in 65-nm technology.

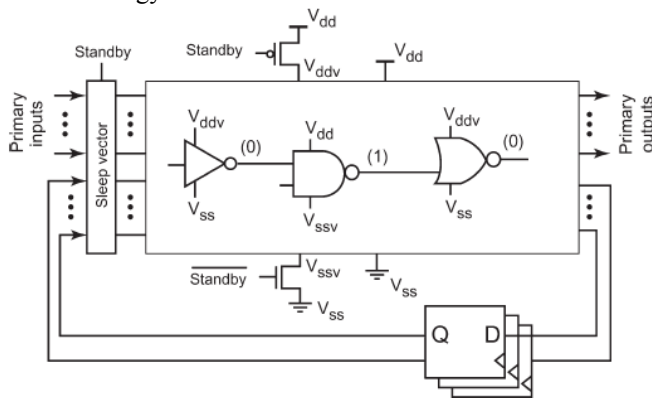


Fig. 8. Zigzag Power Gating (ZPG) Circuit

Liu et al, [26] presented a novel power optimization technique by gate sizing. Gate sizing and threshold voltage ( $V_t$ ) assignment are famous techniques for circuit timing and power optimization. Existing methods are either sensitivity-driven heuristics or based on discretizing continuous optimization solutions. Sensitivity-driven methods are easily trapped in local optima and the discretization may be subject to remarkable errors. In his research, a systematic combinatorial approach is proposed for simultaneous gate sizing and  $V_t$  assignment. The core idea of this technique is joint relaxation and restriction, which employs consistency relaxation and coupled bi-directional solution search. The process of joint relaxation and restriction is performed iteratively to systematically improve solutions. The authors' algorithm is compared with a state-of-the-art previous work on

benchmark circuits. The results from the author's algorithm can lead to about 22% less power dissipation subject to the same timing constraints.

Hyunsik et al [27] developed a technique called variable threshold CMOS, or VTCMOS to reduce standby leakage currents. VTCMOS relies on a triple well process where the device  $V_t$  is dynamically adjusted by biasing the body terminal. By applying maximum reverse biasing during the standby mode, the threshold voltage is shifted higher and the sub-threshold leakage current is reduced. The threshold voltage can be tuned during active mode to optimize performance.

### III. CONCLUSION

As technology scales down below 90 nm, leakage currents have become a critical issue. In the past, circuit techniques and architectures ignored the effects of these currents because they were insignificant compared to the switching currents and threshold voltages were high enough. However, in modern technologies, the role of the leakage currents cannot be ignored and becomes increasingly significant issue with further scaling. Therefore, new circuit techniques and design considerations must be developed to control leakage currents in standby mode in order to provide low-power solutions. After analyzing various leakage reduction techniques, it can be concluded that there is a strong correlation between the three performance metrics: leakage power, dynamic power and propagation delay. If one metric is optimized, it leads to a compromise of other metrics. It can be concluded that super cutoff CMOS scheme provides efficient leakage power savings in standby mode and forced stacking is a very effective leakage power saving scheme for active mode of operation. However, if propagation delay is the main criteria, it is recommended that a single sleep transistor based circuits are used in standby mode, though leakage savings of upto an order of magnitude is sacrificed. In active mode of operation, the sleepy stack based approach is suitable for faster circuit operation.

### REFERENCES

- [1] J. Kao, S. Narendra, A. Chandrakasan, "Subthreshold Leakage Modeling and Reduction Techniques", Proceedings of ICCAD, pp. 141-148, 2002
- [2] Borivoje Nikolic, "Design in the Power Limited Scaling Regime," IEEE Transactions On Electron Devices, vol. 55, no. 1, January 2008.
- [3] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-v Power Supply High Speed Digital Circuit Technology with Multi-threshold Voltage CMOS," IEEE Journal of Solid State Circuits, vol.30, no.8, pp. 847-854, August 1995
- [4] S. Narendra, S. Borkar, V. De, D.Antoniadis, and A. Chandrakasan, "Scaling of Stack effect and its Application for Leakage Reduction," in proceedings of the International Symposium on Low Power Electronics and Design, augusty 2001, pp.195-200.
- [5] Jun Cheol Park and Vincent J. Mooney, "Sleepy Stack Reduction of Leakage Power," IEEE Transactions on Very Large Scale Integration (VLSI) systems, vol. 14, no. 11, November 2006, pp. 1250-1263.
- [6] Preetham Lakshmikanthan and Adrian Nunez, "A Signal Probability based Self-Controlling Leakage Reduction Technique for CMOS Circuits," International Conference on Electrical and Electronics Engineering (ICEEE 2007), September 2007.

- [7] Junbo Yu, Qiang Zhou, Gang Qu and Jinian Bian, "Behavioral Level Dual-  $V_{th}$  Design for Reduced Leakage Power with Thermal Awareness" 2010 EDAA.
- [8] Vivek De, "Leakage-Tolerant Design Techniques for High Performance Processors", in Proceeding of ISPD, Mar. 2002, pp. 28.
- [9] K.S. Khouri and N.K.Jha, "Leakage Power Analysis and Reduction During Behavioral Synthesis", IEEE Transaction on VLSI, 2002
- [10] Xiaoyong Tang, Hai Zhou and Prith Banerjee, "Leakage power optimization with dual-  $V_{th}$  library in high-level synthesis", in Proceeding of Design Automation Conference, 2005.
- [11] Hiroshi Kawaguchi, Ko-ichi Nose, Takayasu Sakurai. "A CMOS Scheme for 0.5V Supply Voltage with Pico-Ampere Standby Current", ISSCC, EEE 1998.
- [12] Lin Yuan and Gang Qu, "A Combined Gate Replacement and Input ector Control Approach for Leakage Current Reduction", 42nd ACM/IEEE Design Automation Conference.
- [13] Behnam Amelifard, Farzan Fallah and Massoud Pedram, "Leakage Minimization of SRAM Cells in a Dual-Vt and Dual-Tox Technology"
- [14] Y. Taur, "CMOS scaling and issues in sub-0.25  $\mu\text{m}$  systems," in Design of High-Performance Microprocessor Circuits, Chandrakasan, W. J. Bowhill, and F. Fox, Eds. Piscataway, NJ: IEEE, 2001, pp. 27–45.
- [15] C. Molina, C. Aliagas, M. Garcia, et al., "Non inferior data cache," in Proc. of International Symposium on Low Power Electronics and Design, 2003, pp. 274-277.
- [16] S. P. Mohanty, R. Velagapudi, and E. Kougianos, "Dual-k versus dual-T technique for gate leakage reduction: a comparative perspective," in Proc. of International Symposium on Quality Electronic Design, 2006, pp. 564-569.
- [17] V. Mukherjee, S. P. Mohanty, and E. Kougianos, "A dual dielectric approach for performance aware gate tunneling reduction in combinational circuits," in Proc. of International Conference on Computer Design, 2005, pp. 431-437.
- [18] R. Chau, S. Datta, M. Doczy, et al., "Gate dielectric scaling for highperformance CMOS: From SiO to high-k," in Proc. of International Workshop on Gate Insulator, 2003, pp. 124–126.
- [19] D. Lee, D. Blaauw, and D. Sylvester, "Gate oxide leakage current analysis and reduction for VLSI circuits," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 12, no. 2, Feb. 2004, pp.155-166.
- [20] Semiconductor Industry Association, International Technology Roadmap for Semiconductors, 2003 edition, <http://public.itrs.net/>.
- [21] Kyung Ki Kim and Yong-Bin Kim, "A Novel Adaptive Design Methodology for Minimum Leakage Power Considering PVT Variations on Nanoscale VLSI Systems," IEEE Transactions on Very Large Scale Integration (VLSI) systems, vol. 17, no. 4, April 2009, pp. 517-528.
- [22] Changbo Long and Lei He, "Distributed Sleep Transistor Network for Power Reduction", DAC 2003.
- [23] Narender Hanchate and N. Ranganathan, "A New Technique for Leakage Reduction in CMOS Circuits using Self-Controlled Stacked Transistors", Proceedings of the 17<sup>th</sup> International Conference on VLSI Design (VLSID'04) 2004 IEEE.
- [24] De-Shiuan Chiou, Shih-Hsin Chen, and Shih-Chieh Chang, "Sleep Transistor Sizing for Leakage Power Minimization Considering Charge Balancing", IEEE transactions on very large scale integration (VLSI) systems, vol. 17, no. 9, September 2009, pp:1330-1334
- [25] Youngsoo Shin, Seungwhun Paik and Hyung-Ock Kim, "Semicustom Design of Zigzag Power-Gated Circuits in Standard Cell Elements", IEEE transactions on computer-aided design of integrated circuits and systems, vol. 28, no. 3, March 2009, pp:327-339
- [26] Yifang Liu and Jiang Hu, "A New Algorithm for Simultaneous Gate Sizing and Threshold Voltage Assignment", IEEE transactions on computer-aided design of integrated circuits and systems, vol. 29, no. 2, February 2010.
- [27] I. Hyunsik, T.Inukai, H. Gomyo, T. Hiramoto, T. Sakurai, "VTCMOS Characteristics and its Optimum Conditions Predicted by a Compact Analytical Model", ISLPED, 2001, pp. 123-128.

# A Survey on Minimizing Energy Consumption of VLSI Processors Using Multiple Supply Voltages

B. Sathiyabama and Dr. S. Malarkkan

**Abstract**—Due to the continuous increase in earth's population, adequate supply of resources is going to be a major issue. One basic essential resource in rising demand is energy and in particular electrical energy. The contributions of the scientific community toward the goal of sustainability with regard to energy consumption of embedded systems are previously discussed in many research works. Low power has become one of the major design issues due to the increased demand in personal computing devices and portable communication system. In this paper a survey on minimizing energy consumption of VLSI Processors using multiple supply voltages is presented. This survey discusses on search method for a scheduling and module selection problem using multiple supply voltages so as to minimize dynamic energy consumption under time and area constraints. The algorithm based on a genetic algorithm is surveyed to find near-optimal solutions in a short time for large-size problems. The literature related to the multiple supply voltages with genetic approach and energy consumption minimization in various VLSI systems is presented.

**Keywords**— Energy minimization, Functional pipelining, Multiple supply voltages, dynamic power, scheduling.

## I. INTRODUCTION

AT the time of writing, 6.7 billion people live on earth. It is estimated that the world population will reach 9 billion by 2040. This increase alone will result in rapidly rising energy consumption. This problem is amplified by rising living standards in developing countries. The present world electricity production is about 20 trillion kilowatt-hours. By 2030, it will reach 30 trillion kilowatt-hours, mostly through coal and natural gas. Coal is a very dirty form of energy, especially when its emissions are not properly filtered, which is mostly the case in developing countries. This will result in serious repercussions for the environment. It stands to reason to try to limit those consequences. One approach is sustainable development, which means to use resources to meet ones needs in such a way that future generations have the ability to meet their needs in the available environment. Producing energy in ways that destroy the environment obviously contradicts this goal. In theory, there is more than enough solar energy available to supply the entire world. In practice, harvesting this energy in a cheap and efficient way is not easy. Therefore it is important to put the created energy to good use instead of wasting it.

The interesting thing to be noted is that the inventor of this device states that this unique power controller design analyses power consumption using an artificial intelligence algorithm implemented on a high-end micro-controller. The key point here is the “high-end micro-controller”, which is a good example for an embedded system and how they can be present without being really noticed, often in surprising quantities. This leads to the main topic, the energy consumption of embedded systems or VLSI Processors and the various strategies available to reduce it. Besides helping to save the environment, reducing energy consumption of embedded systems can lead to immediate monetary rewards for their producers, for example increased sales of the mobile phone with the longest standby time, which is suspected to be the main reason behind the efforts to minimize power consumption. In this survey the Genetic Approach based Minimizing energy consumption of VLSI Processors Using Multiple Supply Voltages is presented.

## II. LITERATURE SURVEY

This section presents the literature survey on the minimization of power consumption in VLSI processors. The power consumption can be reduced only if the cause for the power dissipation is found.

Massoud Pedram [17] presented the cause for the power dissipation in VLSI processors. Power dissipation in CMOS circuits is caused by the three sources: 1) the leakage current which is primarily determined by the fabrication technology, consists of reverse bias current in the parasitic diodes formed between source and drain diffusions and the bulk region in a MOS transistor as well as the subthreshold current that arises from the inversion charge that exists at the gate voltages below the threshold voltage, 2) the short-circuit (rush-through) current which is due to the DC path between the supply rails during output transitions and 3) the charging and discharging of capacitive loads during logic changes.

The diode leakage takes place when a transistor is turned off and another active transistor charges up or down the drain with respect to the first transistor's bulk potential. The ensuing current is proportional to the area of the drain diffusion and the leakage current density. The diode leakage is typically 1 pA for a 1 micro-meter minimum feature size! The subthreshold leakage current for long channel devices increases linearly with the ratio of the channel width over channel length and decreases exponentially with  $V_{GS}-V_t$  where  $V_{GS}$  is the gate bias and  $V_t$  is the threshold voltage. Several hundred millivolts of “off bias” (say, 300-400 mV) typically reduces the subthreshold current to negligible values. With

---

B. Sathiyabama, Research Scholar Sathyabama University, Chennai, India.

Dr. S. Malarkkan, Principal, Manakulavinayagar Institute of Technology, Pondicherry, India.

reduced power supply and device threshold voltages, the sub threshold current will however become more pronounced. In addition, at short channel lengths, the subthreshold current also becomes exponentially dependent on drain voltage instead of being independent of  $V_{DS}$  (see [18] for a analysis). The subthreshold current will remain 102 - 105 times smaller than the "on current" even at submicron device sizes.

The short-circuit (crowbar current) power consumption for an inverter gate is proportional to the gain of the inverter, the cubic power of supply voltage minus device threshold, the input rise/fall time, and the operating frequency [19]. The highest short circuit current flows when there is no load; this current decreases with the load. If gate sizes are selected so that input and output rise/fall times are about equal, the short-circuit power consumption will be less than 15% of the dynamic power consumption. If, however, design for high performance is taken to the extreme where large gates are used to drive relatively small loads, then there will be a stiff penalty in terms of short-circuit power consumption.

The short-circuit and the leakage currents in CMOS circuits can be made small with proper circuit and device design techniques. The dominant source of power dissipation is thus the charging and discharging of the node capacitances (also referred to as the dynamic power dissipation) and is given by:

$$P = 0.5CV_{dd}^2E(sw)f_{clk}$$

where  $C$  is the physical capacitance of the circuit,  $V_{dd}$  is the supply voltage,  $E(sw)$  (referred as the switching activity) is the average number of transitions in the circuit per  $1/f_{clk}$  time, and  $f_{clk}$  is the clock frequency.

Ishikawa et al., [20] discussed on the power dissipation control. A bit-serial multiple-valued reconfigurable VLSI using current-mode logic circuits has been introduced by the author. A Differential-Pair Circuit (DPC) is used as a basic component of a cell, so that the static power is dissipated even in the nonactive cells. To solve the problem, autonomous ON/OFF control of the current sources is presented based on superposition of bit-serial data and current-source control signals. In the proposed switched current control technique, the static power dissipation can be greatly reduced because current sources in nonactive circuit blocks are turned off. The superposition of data and control signals in a single interconnection is effectively utilized to reduce complexity of switches and interconnections, and to eliminate skew between data and control signals. It is evaluated that the reduction of the power dissipation is remarkable, if the operating ratio is less than 75%.

Xin He Al-Kadry et.al, [21] proposed a novel concept to control power dissipation in VLSI processors. This paper emphasizes on adaptive leakage control using body bias technique to reduce the power dissipation of the 65 nm MOS devices. Through adding forward body biasing, the leakage is reduced in sub-100 nm CMOS devices (unlike above-100 nm devices) while slightly increasing the signal propagation delay. For the conditions where the circuit does not use up the entire clock cycle, this slack can be used to reduce the power dissipation without any loss in performance. The fact that the

circuit delay remains less than the clock period provides the opportunity to reduce power consumption of VLSI circuits. The objective is to change the voltage of the body bias to reduce leakage, allowing the circuit to consume less power whenever the clock edge can be met as detected beforehand.

Hariyama et al, [1] proposed a novel approach on minimizing energy consumption of VLSI processors based on dual-supply-voltage assignment and interconnection simplification. His work presents a design technique to minimize energy of both functional units (FUs) and an interconnection network between FUs. To reduce complexity of the interconnection network, data transfers among FUs are classified according to FU types of operations in a data flow graph. The basic idea behind reducing the complexity of interconnection network is that the interconnection resource can be shared among data transfers with the same FU type of a source node and the same FU type of a destination node. Furthermore, an efficient method based on a genetic algorithm is presented for large-size problems.

Xiaoying et al, [25] studied the problem of leakage power reduction by means of input vector control, and develop a platform for CMOS combinational circuit leakage power reduction. Genetic algorithm is worn for searching minimum leakage vector with circuit status difference as fitness function. Experimental results indicate that the proposed method can achieve satisfied leakage power reduction, and the run time is reasonable. This method has no necessity for Spice simulation and independent from target technology.

Hariyama et al, [2] proposed a Genetic approach to minimizing energy consumption of VLSI processors using multiple supply voltages. The author presents an efficient search method for a scheduling and module selection problem using multiple supply voltages so as to minimize dynamic energy consumption under time and area constraints. The proposed algorithm is based on the genetic algorithm so that it can find near-optimal solutions in a short time for large-size problems, n efficient search can be achieved by crossover that prevents generating nonvalid individuals and a local search is also utilized in the algorithm. Experimental results for large-size problems with 1,000 operations exhibits that the proposed method can achieve significant energy reduction up to 50 percent and can find a near-optimal solution in 10 minutes. Conversely, the ILP-based method cannot find any feasible solution in one hour for the large-size problem, even if a state-of-art mathematical programming solver is used.

W. Hung et al, [22] discussed briefly on the techniques in reduction of power consumption. In this paper, the author presents an algorithm for the minimization of total power consumption via multiple  $V_{DD}$  assignment, multiple  $V_{TH}$  assignment, device sizing and stack forcing, while maintaining performance requirements. These four power reduction techniques are correctly encoded in genetic algorithm and evaluated simultaneously. The overhead imposed by insertion of level converters is also taken into account. The effectiveness of each one of power reduction mechanism is verified, as are the combinations of different approaches. Experimental results are provided by the author for a number

of 65 nm benchmark circuits that span typical circuit topologies, including inverter chains, SRAM decoders, multiplier and a 32bit carry adders. From the experimental results, the author shows that the combination of four low power techniques is the effective way to achieve low power budget.

Power dissipation in CMOS digital circuits consists of dynamic power, short circuit power and static power. Short circuit power consumption can be kept within bounds by careful design and tuning the switching characteristics of complementary logic (slope engineering); it is usually negligible compared to dynamic power and leakage power. Dynamic power was once the dominant power consumption term. However, as the result of technology scaling and VTH (threshold voltage) decreasing, leakage power will soon account for a large portion of total power consumption.

Although there are many techniques to reduce power dissipation, most existing works focus on one technique in isolation instead of concurrently applying a number of power minimization techniques. In [22], a power optimization framework based on the genetic algorithm is presented. The optimization strategy combines four power reduction techniques: multiple  $V_{DD}$  assignment, multiple  $V_{TH}$  assignment, gate sizing, and stack forcing. It simultaneously applies and evaluates the effects of these techniques to achieve maximum power saving under a hard timing constraint. The framework can be easily extended to include other power reduction techniques. To the best of our knowledge, this is the first power optimization framework that simultaneously uses all of these four power reduction techniques.

Compared to the ILP approach that were used by [23] [24], one advantage of GA approach is that the parallel nature of genetic algorithms suggests parallel processing as the natural route to explore. The author implements a parallel version of their algorithm, by dividing the population processing among multiple processors. The authors notice that in average more than 3X run-time speed-up on a 4-processor workstation against the single-processor version of the algorithm (The reason that the author cannot achieve a 4X speedup is the interaction overhead among parallel processes). Another advantage is that for ILP approach, the running time for a large circuit may be prohibitively long; while for the GA-based strategy, it can be set a proper termination criterion to tradeoff the runtime and power saving.

Mohanram et al, [3] discussed on the topic of energy consumption in which Lowering power consumption in concurrent checkers via input ordering is presented. The author presents an efficient and scalable technique for lowering power consumption in checkers used for concurrent error detection. The basic idea is to make use of the functional symmetry of concurrent checkers with respect to their inputs, and to order the inputs such that switching activity (and hence power consumption) in the checker is minimized. The inputs of the checker are typically driven by the outputs of the function logic and check symbol generator logic-spatial correlations between these outputs are analyzed to compute an input order that minimizes power consumption. The

diminution in power consumption comes at no additional impact to area or performance and does not require any alteration to the design flow. It is exposed that the number of possible input orders increases exponentially in the number of inputs to the checker. As a result, the computational cost of finding the optimum input order can be very expensive as the number of inputs to the checker increases. The author presents a very effective technique to build a reduced cost function to solve the optimization problem to find a near optimal input order. It scales well with growing number of inputs to the checker, and the computational costs are independent of the complexity of the checker. Experimental results illustrates that a reduction in power consumption of 16% on the average for several types of checkers can be obtained using the proposed technique.

Mohanty et al, [4] puts forth an Energy efficient scheduling for datapath synthesis. In his paper, two new algorithms are described for datapath scheduling which aim at energy reduction while maintaining performance. The given algorithms, time constrained and resource constrained, utilize the concepts of multiple supply voltage and dynamic clocking for energy minimization. In dynamic clocking, the functional units can be worked at different frequencies depending on the computations occurring within the datapath during a given clock cycle. The plan is to schedule high energy units, for instance the multipliers at lower frequencies such that they can be operated at lower voltages to reduce energy consumption and the low energy units, such as adders at higher frequencies, to compensate for speed. The algorithms have been applied to a variety of high level synthesis benchmark circuits under different time and resource constraints. The experimental results demonstrate that for the time constrained algorithm, energy savings in the range of 33-75% are obtained. Similarly, for the resource controlled algorithm, under various resource constraints using two supply voltage levels (5.0 V, 3.3 V) energy savings in the range of 24 - 53% can be obtained.

Muthumala et al., [5] presents a technique to minimize the total energy consumption under time and area constraints, considering interconnection and functional unit energy. Multiple supply and threshold voltage method is used to minimize the static and dynamic energy in the functional units. A genetic algorithm based search technique is proposed for the energy consumption minimization problem, so that near-optimal solution can be found in a reasonable time for large-size problems. Interconnection simplification is attained by increasing the sharing of interconnections among functional units. Experimental results show that up to 30% of energy savings can be achieved by this proposed method.

Kamble et al, [6] presented a comparative study on Energy-efficiency of VLSI caches. The author investigates the use of organizational alternatives that lead to more energy-efficient caches for contemporary microprocessors. Dissipative transitions are likely to be very correlated and skewed in caches, precluding the use of simplistic hit/miss ratio based power dissipation models for accurate power estimations. The authors use a detailed register-level simulator for a typical pipelined CPU and its multi-level caches, and simulate the



execution of the SPECint92 benchmarks to glean accurate transition counts. A detailed dissipation model for CMOS caches is brought up for estimating the energy dissipation based on electrical parameters of a typical circuit implementation and the transition counts collected by simulation. A block buffering method is presented to allow cache energy requirements to be reduced without increasing access latencies. The authors report results for a system with an off-chip L2 cache. The authors conclude that block buffering, with sub-banking to be very effective in reducing energy dissipation in the caches, and in the off-chip I/O pad drivers.

Santosh Chede et al, [7] proposed on the Significance of VLSI Techniques for Low Power Real Time Systems. In microelectronics design, power consumption and the speed of operation, are crucial constraints. Propagation delay of the circuit component has an impact on such factors. Pipelining and parallel processing strategies are used for desirable propagation delays and hence for clock and throughput variation respectively. To some extent variation in propagation delay is accountable for power consumption reduction. In his work, pipelining and parallel processing concepts are analyzed with reference to task scheduling in real time system. Power consumption and speed of operation problems of such systems are analyzed.

Main goal of most of the system level or circuit design are high performance and power optimization. For high performance system design, propagation delay minimization plays an important role. Basically size, cost, performance and power consumption are the crucial issues in low power portable battery operated system design. Excessive power dissipation which overheats thereby degrading the performance and lifetime is not at all affordable. Energy consumption being an important constraint for battery life estimation, VLSI based low power design of dedicated multimode signal conditioning integrated circuit is desirable. Modern systems consist of digital realization of analog processes and this helps to design system with high precision, high signal to noise ratio (SNR), repeatability and flexibility. DSP systems can be realized with custom designed hardware circuits or ultra low power high performance programmable processors fabricated using VLSI circuit technology.

Essentially the role of digital system is to maximize the performance with minimum cost and less time to market. Performance measures are throughput, clock rate, circuit complexity and power dissipation or total energy consumed to execute a real/non real time task. In order to design complex digital system using VLSI technology, modeling with node identification is essential. Generally to carry out design, DSP algorithms are realized and transformed to hardware. To investigate and analyze data flow and data paths i.e. parallelism and pipelining among tasks and subtasks, system modeling methods like block diagrams, Signal flow graph (SFG), Data flow Graph (DFG), Dependence graph etc. is very much required. In such design there is trade off between sampling frequency, operating frequency and power consumption, in order to design high performance system.

Dynamic Voltage Scaling (DVS), Dynamic Frequency Scaling (DFS) can be used to find optimized solution. Various concepts such as pipelining, parallel processing, retiming, unfolding, systolic array etc. are used in design of modern VLSI based low power.

Implementation of VLSI design algorithms includes high level architectural transformations. Pipelining ,parallel processing, retiming ,unfolding, folding and systolic array design methodologies plays an important role for optimized high performance design. Similarly, high level algorithm transformations such as strength reduction look ahead and relaxed look ahead are also utilized for design implementation. Strength reduction transformations are applied to minimize the number of multiplications in convolution, parallel infinite impulse response (FIR) digital filters, discrete cosine transforms (DCTs) and parallel rank –order filters. Look ahead and relaxed look ahead transformations are pertained to design pipelined direct form and lattice recursive digital filters and adaptive filters and parallel recursive digital filters. And these strategies are used to develop and design architectures for multiplication, addition, digital filters, pipelining styles, low power computations and architectures for high performance programmable or ultra low power embedded digital signal processors, applicable to various biomedical, industrial, defense, consumer applications etc [8].

Jui-Ming Chang et al, [9] proposed a novel technique for Energy Minimization Using Multiple Supply Voltages. A dynamic programming technique is presented for solving the multiple supply voltage scheduling problems in both nonpipelined and functionally pipelined data-paths. The scheduling problem refers the assignment of a supply voltage level (selected from a fixed and known number of voltage levels) to each operation in a data flow graph so as to minimize the average energy consumption for given computation time or throughput constraints or both. The energy model is accurate and accounts for input pattern dependencies, re-convergent fanout induced dependencies, and the energy cost of level shifters. Experimental results illustrate that using three supply voltage levels on a number of standard benchmarks, an average energy saving of 40.19% (with a computation time constraint of 1.5 times the critical path delay) can be attained compared to using a single supply voltage level.

One driving factor behind the push for low power design is the growing class of personal computing devices as well as wireless communications and imaging systems that demand high-speed computations and complex functionalities with low power consumption. Another driving factor is that excessive power consumption has become a limiting factor in integrating more transistors on a single chip. Unless power consumption is considerably reduced, the resulting heat will limit the feasible packing and performance of VLSI circuits and systems. The most effective way to lessen power consumption is to lower the supply voltage level for a circuit. Reducing the supply voltage however increases the circuit delay. Chandraskan et al. [10] compensate for the increased delay by shortening critical paths in the data-path using behavioral

transformations such as parallelization or pipelining. The resulting circuit consumes lower average power while meeting the global throughput constraint at the cost of increased circuit area. More recently, the use of multiple supply voltages on the chip is attracting attention. This has the advantage of allowing modules on critical paths to use the highest voltage level (thus meeting the required timing constraints) while allowing modules on noncritical paths to use lower voltages (thus reducing the energy consumption). This method tends to result in smaller area overhead compared to parallel architectures.

There are, however, a number of practical problems that must be overcome before use of multiple supply voltage becomes prevalent. These problems include routing of multiple supply voltage lines, area/delay overhead of required level shifters, and lack of design tools and methodologies for multiple supply voltages. The first issue is an important concern which should be considered by any designer who wants to use multiple supply voltages. That is, there is a tradeoff between lower energy dissipation and higher routing cost. The remaining problems (that is, level shifter cost and lack of tools) are addressed in the authors work. That is, shown that the area/delay overhead of level shifters is relatively small and will present an effective algorithm for using multiple supply voltages during behavioral synthesis.

In this context, an important problem is to assign a supply voltage level (selected from a finite and known number of supply voltage levels) to each operation in a data flow graph (DFG) and schedule various operations so as to minimize the energy consumption under given timing constraints. To this problem is referred as the multiple-voltage scheduling problem or the MVS problem for short. In his work, the problem is tackled in its general form. Chandraskan et al in [10] shows that the MVS problem is NP-hard even when only two points exist on the energy-delay curve for each module (these curves may be different from one module to another), and then propose a dynamic programming approach for solving the problem. This algorithm which has pseudo-polynomial complexity (cf., Section IV-C) produces optimal results for trees, but is suboptimal for general directed acyclic graphs. The dynamic programming technique is then generalized to handle functionally pipelined designs. This is the first time that the use of multiple supply voltages in a functionally pipelined design is considered. A novel revolving schedule for handling these designs is presented by this author.

Bright et al, [11] presents a new tool for the synthesis of low-power VLSI designs, specifically, those designs targeting digital signal processing applications. The synthesis tool genetic algorithm for low-power synthesis represented as GALOPS uses a genetic algorithm to apply power-reducing transformations to high-level signal-processing designs, producing designs that satisfy power requirements as well as timing and area constraints. GALOPS use the problem-specific genetic operators that are specifically tailored to incorporate VLSI-based digital signal processing design knowledge. A number of signal-processing benchmarks are utilized to facilitate the analysis of low-power design tools,

and to aid in the comparison of results. Results show that GALOPS achieves significant power reductions in the presented benchmark designs. In addition, GALOPS generates a family of unique solutions for each design, all of which satisfy the multiple design objectives, providing flexibility to the VLSI designer.

The power consumption of the very large scale integration (VLSI) devices has become an important parameter in recent years, largely due to the explosion in the use of portable communication and computing systems. Of particular interest in such systems are digital signal processing (DSP) devices. These devices are specialized processors used widely in complex functions such as telecommunications, data compression, and speech processing. The increasing requirements for portable systems to incorporate these functions have led to increased demand for low-power DSP devices. However, the design of DSP devices that offer complex functions with low-power consumption requires the use of advanced automated synthesis tools.

Traditional automated synthesis tools optimized the VLSI device for speed and area. This presented a complex solution space, especially when considered at the high level of abstraction (the design capture stage). The addition of power as a design objective compounds the complexity of the high-level synthesis task. Therefore, high-level low-power synthesis tools require a more robust search and optimization mechanism to produce designs with optimal tradeoffs between the objective parameters. Genetic algorithms (GAs) [12] have proven to be a successful technique for tackling the complex problems inherent in the design and optimization of VLSI devices. Examples are VLSI synthesis tools such as that developed by Arslan et al. [13], which uses a GA for the structural synthesis of logic circuits. A tool for reducing VLSI implementation area by using a GA to reduce the size of functional operators was developed by Martin and Knight [14]. The authors have previously demonstrated the application of GAs in low-power synthesis [15], [16] using a restricted GA with limited genetic operators.

In his work, genetic algorithm for low-power synthesis (GALOPS) is presented, a GA for the high-level synthesis of low power CMOS-based DSP designs. GALOP uses problem-specific techniques for low-power synthesis. As illustrated here, the incorporation of these techniques into the GA framework requires modification of the standard genetic operations. The primary objective of GALOPS is to construct a minimum power design from an initial high-level specification, while tracking other performance constraints such as area and speed. The author presents the application of GA for minimizing power consumption in four contributions. (1) The formulation of a GA that can be able in handling the VLSI low-power synthesis problem, considering the specified performance constraints of CMOS devices (2) The development of problem-specific genetic operators that manipulate a library of low-power transformations. Standard genetic operators are modified to incorporate these transformations, in particular, a crossover operator is developed which recognizes and applies power-saving

transformations to designs. The developed crossover operator preserves the notion of inheritance in a genetic algorithm (3) Results that show the significant power reductions obtained using the GA synthesis technique (GALOPS). The effects of relaxing design constraints, such as area, have also been analyzed and compared. (4) Analysis of the capability of a GA-based synthesis tool to present multiple solutions to a problem, exploiting the multiple solution nature of the GA search technique.

### III. CONCLUSION

The reduction in the power consumption is an important issue in modern portable personal computing and wireless communication systems. Minimizing the energy consumption of VLSI Processors using multiple supply voltages with Genetic Approach is a challenge task. Therefore, any method that provides a way to reduce this consumption must be studied, evaluated and applied to the system in development. The techniques reviewed from the literature presented in this survey promise an interesting way to achieve this issue. Also this literature study can be helpful for the further development in the minimization of power consumption in future.

### REFERENCES

- [1] Hariyama, M. Yamadera, S. Kameyama, M., "Minimizing energy consumption of VLSI processors based on dual-supply-voltage assignment and interconnection simplification", Circuits and Systems, 2005. 48th Midwest Symposium on 7-10 Aug. 2005, pp: 1867 - 1870 Vol.2
- [2] Hariyama, M. Aoyama, T. Kameyama, M., "Genetic approach to minimizing energy consumption of VLSI processors using multiple supply voltages", Computers, IEEE Transactions on Issue Date : Jun 2005, Volume : 54 , Issue:6, pp: 642 - 650
- [3] Mohanram, K. Touba, N.A., "Lowering power consumption in concurrent checkers via input ordering", Very Large Scale Integration (VLSI) Systems, IEEE Transactions on Nov. 2004, Volume : 12 , Issue:11, pp: 1234 - 1243
- [4] Mohanty, S.P. Ranganathan, N., "Energy efficient scheduling for datapath synthesis", VLSI Design, 2003. Proceedings. 16th International Conference on 4-8 Jan. 2003, pp: 446 - 451
- [5] Waidyasooriya Hasitha Muthumala, Masanori Hariyama and Michitaka Kameyama, "GA-Based Assignment of Supply and Threshold Voltages and Interconnection Simplification for Low Power VLSI Design", IEEE 2006
- [6] Kamble, M.B. and Ghose, K., "Energy-efficiency of VLSI caches: a comparative study", VLSI Design, 1997. Proceedings., Tenth International Conference on 4-7 Jan 1997, pp: 261 - 267
- [7] Santosh Chede, Kishore Kulat and Rajesh Thakare, "A Significance of VLSI Techniques for Low Power Real Time Systems", 2010 International Journal of Computer Applications (0975 - 8887), Volume 1 - No. 22, pp: 86-91
- [8] Flavius Gruian, "Hard real time scheduling for low energy using stochastic data and DVS processors," in proc. of Int. symposium on Low Power Electronics and Design California, USA, 2001, 46-51.
- [9] Jui-Ming Chang and Massoud Pedram, "Energy Minimization Using Multiple Supply Voltages", IEEE Transactions On Very Large Scale Integration (VLSI) Systems, VOL. 5, NO. 4, DECEMBER 1997, pp: 436-444
- [10] Chandrakasan, M. Potkonjak, J. Rabaey, and R. W. Brodersen, "HYPER-LP: A system for power minimization using architectural transformations," in Proc. IEEE Int. Conf. Computer-Aided Design, 1992.
- [11] Marc S. Bright and Tughrul Arslan, "Synthesis of Low-Power DSP Systems Using a Genetic Algorithm", IEEE Transactions on evolutionary computation, vol. 5, no. 1, February 2001, pp: 27-41
- [12] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Reading, MA: Addison-Wesley, 1989.
- [13] T. Arslan, D. H. Horrocks, and E. Ozdemir, "Structural cell-based VLSI circuit design using a genetic algorithm," in Proc. IEEE Int. Symp. Circuits Syst., vol. 4, 1996, pp. 308-311.
- [14] R. S. Martin and J. P. Knight, "Genetic algorithms for optimization of integrated circuits synthesis," in Proc. 5th Int. Conf. Genetic Algorithms, 1993, pp. 432-438.
- [15] M. S. Bright and T. Arslan, "A genetic framework for the high-level optimization of low power VLSI DSP systems," Electron. Lett., vol. 32, pp. 1150-1151, June 1996.
- [16] Bright, M.S. and Arslan, T., "Transformational-based synthesis of VLSI based DSP systems for low power using a genetic algorithm," in IEEE Int. Symp. Circuits and Systems, CA, May 1998, pp. 45-48.
- [17] Massoud Pedram, "Design Technologies for Low Power VLSI", Encyclopedia of Computer Science and Technology, 1995
- [18] T. A. Fjeldly and M. Shur. " Threshold voltage modeling and the subthreshold regime of operation of short-channel MOSFET's. " IEEE Transactions on Electron Devices, 40(1):137-145, Jan. 1993.
- [19] H. J. M. Veendrick. " Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits. " IEEE Journal of Solid State Circuits, 19(8):468-473, August 1984.
- [20] Ishikawa, Akitaka Okada, Nobuaki Kameyama, Michitaka, " Low-Power Multiple-Valued Reconfigurable VLSI Based on Superposition of Bit-Serial Data and Current-Source Control Signals", Multiple-Valued Logic (ISMVL), 2010 40th IEEE International Symposium on 26-28 May 2010 , pp: 179 - 184
- [21] Xin He Al-Kadry, S. Abdollahi, A., "Adaptive leakage control on body biasing for reducing power consumption in CMOS VLSI circuit", Quality of Electronic Design, 2009. ISQED 2009, pp: 465 - 470
- [22] W. Hung, Y. Xie, N. Vijaykrishnan, M. Kandemir, M. J. Irwin and Y. Tsai, "Total Power Optimization through Simultaneously Multiple-VDD Multiple-VTH Assignment and Device Sizing with Stack Forcing", Proceedings of the 2004 International Symposium on Low Power Electronics and Design (ISLPED'04), pp: 144-149
- [23] Srivastava and D. Sylvester, "Minimizing total power by simultaneous Vdd/Vth assignment," IEEE/ACM Asia-South Pacific Design Automation Conference, pp. 400-403, 2003.
- [24] D. Nguyen, A. Davare, M. Orshansky, D. Chinnery, B. Thompson, K. Keutzer, "Minimization of Dynamic and Static Power Through Joint Assignment of Threshold Voltages and Sizing Optimization," ISLPED 2003.
- [25] Xiaoying Zhao, Jiangfang Yi, Dong Tong and Xu Cheng, "Leakage Power Reduction for CMOS Combinational Circuits", Micro-Processor Research & Development Center, Peking University, Beijing

A.S.A. Kader<sup>c</sup>  
Faculty of Mechanical -Marine Technology  
University Technology Malaysia  
Kuala Terengganu, Malaysia  
abdsaman@fkm.umt.my

paper offer conceptual design architecture for prototyping smart ship towards paperless technology for ships and solutions for barriers that might stand as problem to implementation, innovation and development. This paper focus on building complete macro control system integration between computers, software, hardware electrical, mechanical and other systems as necessary as well as wireless computer network and sensor gateway. The principle that can be use in smart ship shore operations as well redundancy for other smart systems for ship and offshore could target:

- i. Shore base operation : Safe transportation of passengers and cargo through the oceans
- ii. Ship based operations : Loading, unloading, maintenance and servicing of ships
- iii. Management based operations : Cooperative scheduling and controlling of the world's oceans
- iv. Activities for connectivity of the above three

**Keywords-***sytem integration, paperless, ship, eletronics waste*

Technological age has always depends on the concept of speed, reliability, cost, mobility, miniaturization and networking. Not until recently concerned about the environment that has been supporting human and the technocrat that have been built has become the major drive for technology sustainable and efficient system development. Smart ship shore operations focus on building smart system using new information technology and interfacing all the associated service areas of maritime industry - sea and shore. The interconnectivity of control system, computers, software, hardware, electrical, mechanical and other work support system by using in wireless computer network, radio frequency sensor gateway, and the internet for data management activities and operations on board ship. The

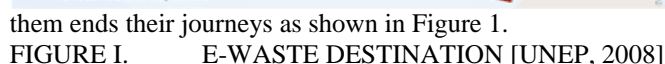


TABLE I. COMPONENTS OF COMPUTER

Material	%	-
Ferrous metal	32	-
Plastic	23	-
Non ferrous metal	18	Lead cadmium antimony beryllium, mercury
Glass	15	
Electronics board	12	Gold silver platinum

In the contrary, the computer help us to reduce solid waste of papers, and a thorough cost benefit analysis is required to justify use make balance between use of computer and the computers disposal at the end of their life. How nice it will be to have semi conductor industry to recycle the use of the material. Another problem is the effects of the trace metals on human health and ecology. Tailoring IT utilization list for a specific ship application and associated seaport facility, passenger and cargo carrier, or sea traffic management systems require reference, networking interfacing and integration of infrastructure worksheet of the list of core technology elements associated with system support resources associated services in ship shore operations like : Administration, Finance, Building, including Access, Environmental, Inspection, Security and Safety as well as Communications, Office Equipment, Computer Network Resources, Food Storage and Distribution, Interface, Public Works, Software, Emergency Medical Systems, Marine Systems, Rail and land intermodal system linked with the port operations and door to door services. Necessary interface and integration software could be adapted to facilitate this intelligent transportation work as required.

## II. REQUIREMENT FOR SHIPBOARD PAPERLESS SMART SYSTEM

Today things that have been developed by human are becoming faster and the time to do many, to finish or do more things is not enough, give rise to today's need for systems integration. In other to further build capacity for use of ICT, and boost its contribution to mitigation of climate change it is important to implement the efficient and sustainable smart system for ship this include [1]:

- Design of universal control server for shipboard power control systems.
- Building of artificial intelligent development program for the system - With a result to reduce man power, simplify management process and cost reduction.
- Connection of the central monitoring station to wireless node, sensor network including the internet

through fiber optics and blue tooth -this for easy data transfer and control needed.

- Building of a dynamic intelligent system that will provide easy synchronization to automated computation in shipping industry.
- Building of scalable and efficient system that will reduce transportation cost, ease system management and upgrade

Other benefits that can be derived from such system are:

- Production of potential manpower savings while improving readiness and quality of life through assessing policy changes, innovative cultural and tradition modifications and integration of advanced technology to come up with the need of cybernetic ships.
- Reduction of annual operating costs by million, and developed long-term benefits in terms of improved morale and retention of quality personnel. Multimedia digital system on board ship, creation of more space from recreation and training of crews that can will improve efficiency, reduce homesickness of the crew and improve quality of life for the crew.
- Increase in efficiency of the shore infrastructure, QOL increases for operational forces through better customer service and for installation personnel through reduced workload.
- Increases in efficiency and subsequent increases in QOL and increase operational readiness
- Contribution to absolute safe monitoring and management of utility power plant for high reliability, black out detection, and emergency cut out efficiency
- Low operating cost and reduced human power as well as -E-billing facilitation and e- interface capability
- Stand alone and remote control capability from a long range.

## III. SCOPE OF SUSTAINABLE SMART PAPERLESS SHIP PROJECT COULD LIE WITHIN THE FOLLOWING

- **Shipboard operations** - Alarm Data loggers, Bridge Movement Recorders, Compass/Internal Navigation, Commercial Radar, Doppler Speed Logs and accessories, Electronic chart systems, - electronic charts, bridge systems and equipment, automatic identification systems, shipboard alert systems, also changing role of class societies, regulation and policy thematic areas [2].
- **Port terminal** - Elevators, Escalators, Fire Detection, Heating and Air Conditioning (HVAC), Electric Power, Safety, Security, Telecommunications, Water

and Sewer systems, as well as Leasing and Financial Management Software to support the Building Infrastructure that required to support the various aspects of the Sea Transportation activities.

- Integration- Integration and interface through smart and wireless network.
- **Software** - Word Processing software ,Spread Sheet software for various reporting system ,Briefing software , Scheduling software, Data base software Electronic Mail software to exchange informal correspondence, Network Browser software, shipboard e-mail software, maintenance management, computer based training, simulators, web services, shipping company software, communications with brokers, charterers and agents, shipping agent systems .
- **Hardware** - Computer Hardware used to run the software to satisfy the Sea Transportation sectors daily automated IT requirements, Hardware used to manage, monitor, and control radar and other ship management aids,
- **Communication** - Communications systems, both voice and data, used to coordinate and facilitate the flow of ships, on-ground vehicle traffic, and requests for seaport services. Inmarsat Fleet, shipboard information services, ship tracking, ship-shore communications by e-mail, data and voice, Iridium, sending weather, chart updates, training materials to ship, sending ship operations information back to shore, Fax, photocopiers, telephones, mobile telephones and radios, vehicles, etc. Global Positioning System (GPS), Radar, Periscope Personnel computers and associated software Steering Vessel Measuring System (VMS) and Vessel Traffic System (VTS)
- **Security** – The facilitation of the of the use smart cards to replace our various conventional security systems. Practical application of the ISPS code, training and monitoring the ISPS code, seafarer ID cards, reducing shipboard piracy, container shipping security, port security, Smart and Secure Trade lanes, and container security initiatives.

#### IV. PROCESS FOR SHORE SENSOR INFORMATION PROCESSING SYSTEM (SSIPS)

To put the SSIPS system together, the process could target the following three phase viz:

- i. Equipments and testing,
- ii. Arrangement, connection, networking and testing
- iii. Coding and testing.

A *Description of Equipment*

1 *Hardwires*

- i. Main devices: PDA, A/D converter, microcontroller device, pc
- ii. Network sensors-infrared remote control, water level controller, RS 232 interfaces for A/D converter, digital automation temperature control smoke detector fire alarm controller.
- iii. Testing equipment: digital voltmeter, 5volt power supply, digital logic probe, 4 channel digital scope or universal Tektronix logic scope and analyzer.

2 *Software*

- i. Initial design and simulation software: CAD, MathCAD, electronics workbench will be use for initial design, Simulation software, Atic-transmission line simulation, Aviprogram-microcontroller device in parallel programming ,Digitemp-read temperature data from USB or wire, Eep- read and EPROM devices
- ii. Computer language: assembly language for microprocessor communication
- iii. Network integration softwares java and c++ for wireless LAN communication between the PDA, the device and the sensor and visual basic for auxiliary Communication need and solution

#### V. GENERAL SSIPS ARCHITECTURE

Wireless device with capability to use LAN and sensor network need to be established as needed. Analogue to digital converter will be wired to a computer .The PC could be the main control center that will be loaded with hardware and software interface program needed. This will be the main control station through with features can be changed and tested on the devices. Network sensor and Bluetooth and connection to various systems can be established to the various computers, wireless device and sensor and the internet. The layers of the architectures include[3,4]:

- Server layer: Micro servers are used to host local content and may be embedded in various devices to make them network-aware and remotely controllable. Can be used TINI embedded microcontroller to build the micro server TINI is the size of a 72-pin memory SIMM module, has a hybrid 8/32 CPU (DS80C390 - backwards compatible with 8051), 512K/1M non-volatile (battery-backed) RAM, 512K flash ROM, 2x serial, 2x CAN, 1-wire, parallel, real-time clock, and runs Slush (with a command shell similar to Unix), a Java virtual machine (JVM) and Java API. Software includes FTP server, TELNET server, TTY server and HTTP server. A micro server may have interfaces such as CF (mainly for networking), SD (mainly for storage), USB, infrared, and DSP and Clock speed in the order of 300- 500 MHz.



- **Wireless layer:** For a user can interact with the cyber-assisted environment through an agent device like PDA's or mobile phone. Wireless network with interoperability being one of the most important aspects, the standard 802.11b and 802.11 will be used according to nature of the system.
- **Sensor network layer:** networks are used to get real world data, here there is need for gateway to IP sensor, The main module has a microprocessor and an RF chip; the sensor module has the sensors for light, temperature, magnetism, acceleration, and sound. As the role of sensor networks is constant monitoring, radio communication is usually in use. Therefore, gateways between sensor networks and computer networks will be needed; their task is to monitor incoming sensor data and invoke hook operations when trigger events are detected.
- **Middleware layer:** To have applications working over heterogeneous embedded networks requires lightweight mechanisms of service directory and remote procedure invocation. Multicast (DNS) domain name service discovery is a good pick for the former functionality and the web based technologies for the latter, as JAVA for simple server-side processing and SOAP (simple object access protocol) for network wide distributed processing.
- **Real world user interface:** Implementation of smart ship will allow user to interact with the real world environment, user interface on the agent device will be a good in thing in addition to the standard web browser interface. Infrared communication will be used for the physical reference, where the IP address of the agent device and the URL of the service on the micro server is pointed at by the user are exchanged. Voice recognition will be useful for specifying services, where the user's voice is sampled on the agent device and then transmitted to the voice recognition server by the VoIP (voice over IP) technology. The result may be used with the multicast DNS service discovery. Another interesting method is location-based service lookup. Other location infrastructures that can be used are 3D video tracking and based station id of wireless will be used for short distance. Switch arrangement to GPS should be provided for ship-shore needs.
- **Security layer:** the security and privacy issues are ever prominent yet potential difficult and problems in smart system design. It is not acceptable for the ship or operations to fail. And since the whole idea is about opening the network to the net, securing the network is important; security measure will include ID encryptions and using location authentication. Using internet security for protection against long distance intruder.

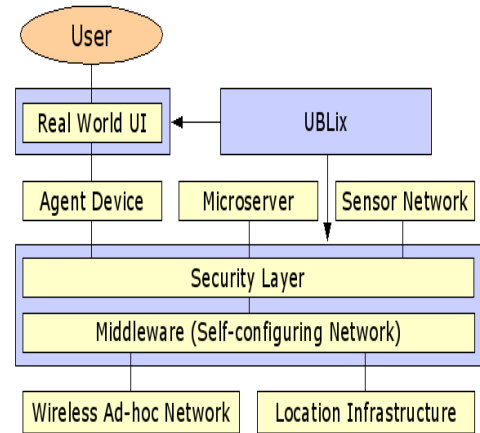


FIGURE II. SYSTEM ARRANGEMENT AND DIAGRAM

## VI SUBSYSTEM OPERATIONAL FRAMEWORK

This system can be tested on different subsystem and system, technology could be applies on various part of the ship: Most ship systems are design to run on analogue built system, in this new digital time, today there are many commercial and ship built on digital system. Integrated Ship Control System (ISCS or ISC) is replacing obsolete analog ship's control systems with a state-of-the-art, commercial off-the-shelf (COTS) system with a Windows NT operating system or Linux that communicates via dual-honed single and multi-mode fiber optics. The air-blown fiber optic network, programmable logic controllers (PLCs), data acquisition units (DAUs), Intergraph computers, and Henshel hardware components are the backbone of the equipment equipments in this category [5].

## VII BEST PRACTICE FOR SHIPBOARD SYSTEM

### A. Typical example for on board integrated Bridge System (IBS)

This include provision of navigation assurance via automated piloting and ship's course and track analysis with radar and digital nautical chart overlay, as well as collision avoidance, on possible required integration needed from on board ships systems are:

1. *Actuators*
  - i. 4 rpm controlled azimuth thrusters with indecently controllable azimuth angles that can control commands are transferred from the control room to the ship by a radio link.
2. *Computer system*

- ii. Micro PC onboard the ship running QNX real-time operating system (target PC)
- iii. The control system is developed on a PC in the control room (host PC) Simulink/Opal can be used on target PC using automatic C-code generation and wireless ethernet.
- iv. Integrated Condition Assessment System (ICAS) that provides condition-based maintenance recording for main propulsion, electric plant, and auxiliary equipment
- v. Wireless Communication System that provides handheld communications for ship's key personnel in or near the ship.

*B Typical example - for a on board power plant and fuel control system*

Machinery Control System (MCS) could provide main propulsion and electrical plant control, also fuel Control System (FCS) provides an automated control of a ship's fuel fill and transfer operations.

- i. Determination of power capacity of a particular utility plant and how many substations it can be divided into.
- ii. Number of equipment will be derivative of 1
- iii. Number of substations according to power need with step up and step down transformers will be ideal for a prototype design. With nominal 220v input and output voltage.
- iv. 4-Auto transformer can be used to determine input and impedance can be used to simulate the load and filters will be connected for PF control station, substation central station connected through LAN and Bluetooth to the various computers, wireless device and sensor and the Internet.
- v. Number of computer-aided system can be connected to the switches and controller and various network devices, but whenever it cannot find a solution it will send alarm or message through the network to human and will learn how the solution was provided.
- vi. 3-phase transformer Current transformer Potential transformer Autotransformer Impedance, Circuit breaker, Computers fixed filter and Control filters Switch software simulation.

Focal areas for this model are intelligent routing of the system and Power distribution and control. The use of existing power line to build the network physical and control could enter e-recycling regime.

*C. Typical example for smart terminal operations*

This will include capability for short range sensing application ashore and on board where the system allows the user to control any device that has ordinary IR remote

controls. There is need to prepare a micro server with an infrared port, run a web server with a java script that has information about the remote control signals for the target device, and placed it in an appropriate position. The remote control procedure will be according to following steps could follow:

- i. The user transmits the agent device's IP address in the infrared beam by pressing a designated button. The micro server receives the beam and sends back the URL of the remote control java page on its web server to the decoded address of the sender over the wireless network.
- ii. A web browser is started with the returned URL on the agent device; the user then sees a remote control page for which he/she presses one button. A request is sent to the micro server's web server; next, the corresponding JAVA script is executed and the control signal is transmitted to the device. The user sees his request granted. Voice IP technology will be incorporated as necessarily

*D. Typical example for ship shore communication and data sharing*

This involves we will develop application to open and integrate ship shore network and communication system to the internet. This way, communication will be cheaper, but the thing here is that main security satellite transmission and reception dish will be used here.

## VIII WHAT CAN BUILDING THE ABOVE DESCRIBE SYSTEMS ACHIEVE

*A. Paperless desktop*

- i. Regional information/data repository
- ii. Single interface for all Web-accessible information/applications
- iii. Automatic routing and storage of correspondence and instructions
- iv. Reduction in labor associated with copying and mailing paperwork
- v. Faster access to more accurate information increases worker satisfaction
- vi. Reduced workload for admin workers provides more time for other essential tasks

**\*\* Naval Air Station (NAS) Brunswick: saves 36,000 sheets of paper annually on instruction distribution alone**

*B. Smart terminal*

- i. Automated port operations management system

- ii. Vessel/Berth/Tug Scheduling & Web Enabled System
- iii. Provides improved efficiencies by providing improved communications and near "real-time" monitoring of Port Ops assets
- iv. Reduces preventive and corrective maintenance on all port operations equipment, which means less overtime required for already-busy sailors and civilians

*D. Smart Procurement Electronic Data Interchange (SPEDI) has been implemented in the US NAVY ship*

- Total e-commerce solution for negotiated procurement
- Automated back office processing capability and interface to DFAS
- Completely eliminates unmatched disbursements
- Just-in-Time/optimized inventory management
- Reduces labor associated with order, receipt and bill paying procedures (customers, vendors)

*E. Activity-Based Costing/Management (ABC-ABM)*

- Benefit derived here will be
- Provides tools to facilitate management decisions in allocating resources efficiently and maximizing operational efficiencies
- Operational efficiencies gained reduce maintenance requirements, which in turn reduce workload
- Savings gained through process reengineering can be used to better serve the sailor

**F. Intransit material bar-coding**

- Automated inventory tracking and notification system
- Distribution Center automatically notifies customers via email of supply shipments
- More accurate tracking of supplies
- Faster delivery of supplies from distribution center to operational units
- Supplies directly or indirectly affect QOL of sailor (operational readiness)

*G. Electronics Security*

- Automated electronic access control via Smart Card using contactless card technology (proximity card)
- Monitors general base access and special access areas
- Central secure surveillance of interiors, exteriors and property parameters
- Records management system/access audit/ event recording

- Provides increased security status visibility
- Reduces manpower requirements while increasing security level
- Increased security means safer working environment

*H. Lifeline*

- Delivers QOL services and programs via the Internet, teleconferencing, satellite broadcasting and cable TV
- Web site uses "expert systems" to deliver a comprehensive range of human services and assistance to the total force
- Assurance of supplements does not replace traditional QOL community-based service delivery systems
- Provides greater access to "high touch" human services using "high tech" modern technologies
- Maritime multimedia super-corridor-including training video conferencing  
Joint maritime services partnership

*I. Smart base direction*

- Near Term:
  - Installation and Region Interoperability
  - Local Operational Unit Integration into Shore Business Process
- Mid-Term:
  - Multi-Region Interoperability
- Long Term:
  - Installation, Region, Headquarters Staff Interoperability
- Continuous:
  - Leverage existing investments, infrastructure, policies and practices to assist in reducing the cost of daily shore-based operations

**XI COMPARATIVE ADVANTAGE**

Theoretical knowledge based on digital electronics, LAN, microprocessor, c++ and other electrical, electronics, computer and communication engineering and ship systems areas could be applied as needed. Installations of systems that perform voyage management, digital damage control information management, wireless internal voice communications, equipment condition assessment, digital machinery control, a Learning Resource Center and a fiber-optic LAN is recommended for the Smart Ship System Assessment report. This is required utilize the advantage of the synergy between technology and policy and procedural changes and to enable manpower reductions with less risk.

Risk solutions and recommendation –the military has always been the first institution that invest a lot to try new things, the US Navy's has been involved in smart ship program geared towards reduced cost and man power, while also increasing efficiency. Installation and integration of the

latest technologies now provides the fleet with automatic digital control and status monitoring for its vital operational systems thus this has always been the trend of research work, adapting current system developed smart system on war ship and smart terminal operations to commercial fleet is a presently needed for the level of technology in shipping.

The cost of establishing this system and the network is high - however; suggestion will be made on upgrade of existing facility to lower cost to a greater degree. The above proposed describe system is a complex task, since two heads are better than one combine industry, military, academic and professional in this field has been useful and this could facilitate the other work including building the prototypes , system and equipment installation.

Survey and analysis of existing wireless and network system and modification to requirement is a good tool to identify various goal and necessary upgrade, which will equally reduce the cost requirement of connection and equipment that will be employed for hybrid facility Selection and building the program tool to execute the control, operation, and research will be made on existing related program in other field as well as code elaboration will be developed.

#### X Cost and benefit analysis

The cost of analysis to complete the control and network system requires critical analysis accordingly. The long term objective of these projects is to identify the most promising labor saving technologies available today for back fit, or projected the future for forward fit. This will potentially save the industry significant manpower funds, in the long term. Figure III shows cost saving from

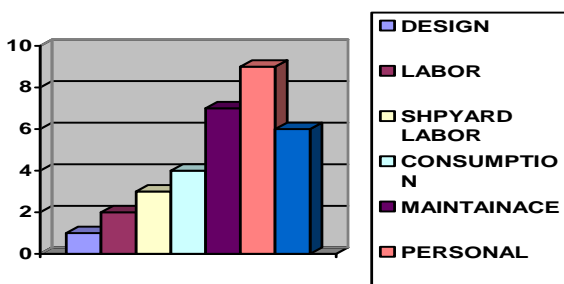


FIGURE III. TOTAL OWNERSHIP COST COMPONENTS FOR A SHIP [SOURCE: NAS, 2009]

Traditional ships need a lot personal, and cutting out this number has always been a priority for many managers in the transportation industry and the answer to this is applications of information technology challenge. It will avoid Duplication or overlap with Smart Ship endeavors by working with other industry and merge ideas together to come up with the best aim system to control our devices., except in those cases where the initiative is an element of the core

infrastructure are necessary to implement modern technologies. Table II shows potential ship board impact areas.

TABLE 11. SPACE WHERE SIMILAR POLICY CHANGE COULD IMPACT

Space	Policy Impact	Application
Ship office	Paperless ship reduced, and reduction in weight of paper.	Potential impact on need for separate Log Room, engineer's office, etc.
Electronic logbook	Operating logs pared down, periodicities reviewed and revised and data log entry automated portable data terminals,	Employment of automatic sensors and a data base system
Pilot House	Reduction of watch	Use of Ship's Control Console
Automation	Reduction of watches by removing EW Supervisor, plotters, etc.	Use of automated multi-purpose consoles. Reliance on automated systems will require accelerated system rest oral times after casualty
Quarterdeck	Reduced to single watch	Simplified interior communications and alarm access. Automated deck log capability

Information and initiatives will be freely exchanged and shared to increase the efficiency and effectiveness of both programs could augment the following:

- Technology transfer prospect: in a world that is going through a wireless revolution, product reduction, reliability, scalability and speed, the above described. Smart system will be very marketable and its patents will allow technology and transfer and cooperation within electronics and marine industry. Same redundant technology can be use for smart hem and smart industry and facility management.
- The advantages of the proposed design would reduced manpower, reduced capital investment, reduced demand for hot/fresh water, improved sanitation, individualized menu selection, no inventory requirement, and reduced storage space. Issue of Intelligent electric power resource management and other electrical or electronic power management are ample areas for improvement. This technology, application to other platforms, will shown savings in operational and maintenance time.
- Potential Improvement: Independent of definitive billet reduction, impacts on design may be measured. Obviously, spaces should be designed to work efficiently, whether manned to historical levels or revised to reflect improvements in technology or

policy revision. Equipments would be multi-functional or accessible to more than one operator. Physical area may be reduced as appropriate to facilitate access by fewer personnel. Efficiencies in stowage are also examples of where design can be modified to best serve the crew regardless of size. On Smart Ship, the majority of the workload savings can be attributed to policy and procedure changes. A new mind-set to review business as usual and develop methods to utilize only those personnel required to guide this effort [6, 7].

flow management, and personnel/facilities scheduling.

Sustainability through manning and shipbuilding design:

- i. Manning through constant care can only be maintain through training of personal, establishment ICT department to focus on improving ship's performance on board, this project will be beneficial to the improvement of work on ship and shore. First, shore-based training is transitioning to shipboard responsibility - and secondly, there will need for embedded on board training program with emphasize on expanded computer skills for both operator and maintainer and average sailor's tenth grade educational level has to be supplemented by extensive onboard training.
- ii. Shipbuilding design based on the above will be a good guide for ship designer, their space work and partition. Other recommendations are:
  - a. Fleet Modernization and conversion - Full build integrated design, logistical and planning and management support for modernize old fleet.
  - b. Upgrading fleet with Advanced Planar Antennae - use of Integrated VHF/UHF/L-Band Antenna (IVUL), Advanced Multifunction Radar Frequency Concept (AMRF-C).
  - c. Radar upgrade - Ship systems to use near field radar reflectivity range that provides the capability to measure the radar reflectivity of large structures accurately and with high resolution. Individual component contributions.
  - d. Remote Source Lighting - remote source lighting technology in marine applications that allows full spectrum effective lighting in hazardous atmosphere environments without the expensive, heavy weight explosion.
  - e. Fiber optics cable technology promised weight reduction and lowered interference susceptibility for communications applications.

## XI RISK AREAS FOR SSIPS

**Major Potential Risk Areas :** for everything the is always the bad side, smart system development for ship has no exemption , therefore extensive work on security for necessary protection.

- i. Potential Area(s) of Automation Risk: building infrastructure, office and Network systems, and the core seaport/terminal, ship carrier, and ship traffic management Infrastructure could affect the sea transportation sector's ability to conduct critical, essential, and non-essential business. The inability to utilize mobile radios, mobile telephones, telephones, and utility vehicles including support systems e.g., maintenance, etc., and non-essential systems impact would include the inability to utilize basic office support resources
- ii. Operational: Loss of failure of the vessel's navigation, steering, speed logs, military and commercial radar, Vessel Measuring System (VMS), Vessel Traffic System (VTS), communications and Periscope, control systems, e.g., engine and cargo systems, failure or corruption of the seaport, sea carrier, or sea traffic management systems. Data Bases that support the fundamental sector operations such as passenger and cargo reservations and booking, sea planning and traffic flow management, and personnel/facilities scheduling, Mobile Radios and Telephones, Utility Vehicles, and Telephone Service
- iii. Port Infrastructure: Loss of elevators, escalators, fire detection, heating and air Conditioning (HVAC), electric power, safety, security, telecommunications, reservation, flight planning, and scheduling systems, as well as Leasing and Financial Management Software. Failure or corruption of the seaport, sea carrier, or sea traffic management systems. Also, data Bases that support the fundamental sector operations such as passenger and cargo reservations and booking, customer billing, sea planning and traffic

## XII MITIGATION OPTIONS ON IMPACTS

- i. Officer of the Watch manned only when as necessary, change may not have negative impact on readiness and can be implemented at zero cost. Wireless communications provides flexibility to have these personnel roving or on-call rather than remaining in concentrated on job. Design sufficient monitoring, Issue: Shipboard Automated Planned Maintenance System (SAPMS) reduced time to perform PMS checks on UHF equipments Impact. SAPMS or similar system would be of more benefit because of potentially more use of UHF equipments

- ii. Use of High Speed Fleet Broadcast system will increase the throughput rate of fleet broadcast by 300% to 9600 bps and enhances ship's HF capabilities by providing single tone modems with interleaving and Forward Error Correction. It replaces existing fleet satellite broadcast modems. Message volumes continue to increase and even, hence higher speeds will be needed. Modems already achieving 56.6K bps a faster system will be beneficial.
- iii. High Frequency Radio Group replaces existing shipboard manually tuned/controllable and rigid HF (2-30MHz) systems. HF remains a viable tactical circuit for ship-to-shore operations particularly due to the always present competition for satellite bandwidth. Automating tuning and controlling will increase accuracy and set up speed.
- iv. Touchpads replace mice and/or track balls facilitate operator interface and have no moving parts. There may be difficult for the operator because the pitch size of the indicator will be too small to see under the finger and the sensor areas had to be brightly lighted in order to find - impacting night vision. Suggestion: Equipments with Touchpad should be evaluated by location, requirement for lighting, and ease of using before full scale implementation
- v. EDI system- Resale Operations Management Electronic Data Interchange (ROM EDI) eliminates various paper-intensive operations such as preparing hard copy, monthly transmittals. It improves accuracy while reducing workload.
- vi. Automated Log Keeping, reduce paper work on board ship and improve accuracy.
- vii. Galley maintenance- is a labor intensive function. It is possible to modify the bulk of this operation without overburdening the individual Sailor through innovative packaging and food preparation. The proposed conceptual design would require an automated ID system, a comprehensive data base, advanced food processing and packaging, and innovative heating technology.

### XIII CONCLUSION

Beside environmental benefit of reducing flooding of land and capturing of green house gases by less tree cutting,

smart ship technology will achieve a return on investment in future voyage management, LAN/wireless internal communication and machinery condition assessment that are valuable particularly in reduce risk, and increasing speed. Smart new technology applications will supplement or replace current shipboard systems and subsequently mitigate increased risk perceived from reducing watch standers. SPS could also reduce present day's transportation problems in making transportation system safer and more efficient. Setting, program functioning, and staff skills unification of system in diversity, foster extensive cooperation among professionals, agencies through building integrated information system. It will also improve readiness for change and resources available on pressing issues in marine transportation systems as well further great transformation and reciprocating performance contribution to human civilization.

### REFERENCE

1. Sumit Ghosh Tony Lee, 2000. Intelligent transportation. modeling and simulation techniques for ITS Design, future issues of ITS,CRC ,PP135-165
2. C.S. Papacostas, P.D. Prevedouros. Transportation engineering and planning, prentice hall 2001.
3. Roshini Ratnasingham, Indept study of oxide material for high frequencyapplication  
<http://www.sysopt.com/articles/soi/index3.html>
4. <http://www.cs.berkeley.edu/~pal/research/tossim.html>, accessed, may, 2009.
5. Dan Houser. Infosecurity magazine, network security for submarines, pp46-57, august 2003.
6. Geographical information system in transport research. Pergamon 2000.
7. Mechatronics.W.Bolton. Prentice Hall, 2003.

### AUTHORS PROFILE

- O. O. Sulaiman is currently lecturer and research fellow in Maritime Technology department at University Malaysia Terengganu. He chattered engineer and chattered marine enegineer. He has 18 years experience in maritime industry. He has taught various maritime subject. He has sailed on board ocean going vessel vessel and work on various marine projects.
- A.H. Saharuddin is curently the dean of of Faculty of Mraitime Studies and Marine Science. He is a Fellow of IMarEST. He is a policy expert in maritime industry.
- A.B. S. Kader is profesor of Marine Technology, He has worked on various maritime projects.



## IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA  
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia  
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA  
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway  
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India  
Dr. Amogh Kavimandan, The Mathworks Inc., USA  
Dr. Ramasamy Mariappan, Vinayaka Missions University, India  
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China  
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA  
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico  
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India  
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania  
Dr. Junjie Peng, Shanghai University, P. R. China  
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia  
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India  
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain  
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India  
Mrs Li Fang, Nanyang Technological University, Singapore  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia  
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India  
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand  
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India  
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia  
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India  
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India  
Mr. P. Vasant, University Technology Petronas, Malaysia  
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea  
Mr. Praveen Ranjan Srivastava, BITS PILANI, India  
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong  
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia  
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan  
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria  
Dr. Riktesh Srivastava, Skyline University, UAE  
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia  
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt  
and Department of Computer science, Taif University, Saudi Arabia

Mr. Tirthankar Gayen, IIT Kharagpur, India  
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan  
Prof. Ning Xu, Wuhan University of Technology, China  
Mr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen  
& Universiti Teknologi Malaysia, Malaysia.  
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India  
Mr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan  
Prof. Syed S. Rizvi, University of Bridgeport, USA  
Mr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan  
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India  
Mr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal  
Mr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P  
Dr. Poonam Garg, Institute of Management Technology, India  
Mr. S. Mehta, Inha University, Korea  
Mr. Dilip Kumar S.M, University Visvesvaraya College of Engineering (UVCE), Bangalore University,  
Bangalore  
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan  
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University  
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia  
Mr. Saqib Saeed, University of Siegen, Germany  
Mr. Pavan Kumar Gorakavi, IPMA-USA [YC]  
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt  
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India  
Mrs.J.Komala Lakshmi, SNR Sons College, Computer Science, India  
Mr. Muhammad Sohail, KUST, Pakistan  
Dr. Manjaiah D.H, Mangalore University, India  
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India  
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada  
Dr. Deepak Laxmi Narasimha, Faculty of Computer Science and Information Technology, University of  
Malaya, Malaysia  
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India  
Mr. M. Azath, Anna University, India  
Mr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh  
Mr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia  
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore  
(MP) India,  
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia  
Mr. Hanumanthappa. J. University of Mysore, India  
Mr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)  
Mr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria

Mr. Santosh K. Pandey, Department of Information Technology, The Institute of Chartered Accountants of India

Dr. P. Vasant, Power Control Optimization, Malaysia

Dr. Petr Ivankov, Automatika - S, Russian Federation

Dr. Utkarsh Seetha, Data Infosys Limited, India

Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal

Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore

Assist. Prof. A. Neela madheswari, Anna university, India

Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India

Mr. Kamanashis Biswas, Daffodil International University, Bangladesh

Dr. Atul Gonsai, Saurashtra University, Gujarat, India

Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand

Mrs. G. Nalini Priya, Anna University, Chennai

Dr. P. Subashini, Avinashilingam University for Women, India

Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat

Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal

Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India

Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India

Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah

Mr. Nitin Bhatia, DAV College, India

Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India

Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia

Assist. Prof. Sonal Chawla, Panjab University, India

Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia

Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India

Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France

Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India

Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban,South Africa

Prof. Mydhili K Nair, M S Ramaiah Institute of Technology(M.S.R.I.T), Affiliated to Visweswaraiah Technological University, Bangalore, India

M. Prabu, Adhiyamaan College of Engineering/Anna University, India

Mr. Swakkhar Shatabda, Department of Computer Science and Engineering, United International University, Bangladesh

Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan

Mr. H. Abdul Shabeer, I-Nautix Technologies,Chennai, India

Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India

Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India  
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran  
Mr. Zeashan Hameed Khan, : Université de Grenoble, France  
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow  
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria  
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India  
Dr. Maslin Masrom, University Technology Malaysia, Malaysia  
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India  
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City  
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE  
Dr. Abdul Aziz, University of Central Punjab, Pakistan  
Mr. Karan Singh, Gautam Budtha University, India  
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India  
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia  
Assistant Prof. Yasser M. Alginahi, College of Computer Science and Engineering, Taibah University, Madinah Munawwarah, KSA  
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India  
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India  
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India  
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India  
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India  
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia  
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India  
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India  
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius  
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India  
Dr. Mana Mohammed, University of Tlemcen, Algeria  
Prof. Jatinder Singh, Universal Institutiion of Engg. & Tech. CHD, India  
Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim  
Dr. Bin Guo, Institute Telecom SudParis, France  
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius  
Prof. Pijush Biswas, RCC Institute of Information Technology, India  
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia  
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia  
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius  
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore  
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India  
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India  
Dr. C. Arun, Anna University, India

Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India  
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran  
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology  
Subhabrata Barman, Haldia Institute of Technology, West Bengal  
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan  
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India  
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India  
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand  
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India  
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.  
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran  
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India  
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA  
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India  
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India  
Mr. Serguei A. Mokhov, Concordia University, Canada  
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia  
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India  
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA  
Dr. S. Karthik, SNS College of Technology, India  
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain  
Mr. A.D.Potgantwar, Pune University, India  
Dr. Himanshu Aggarwal, Punjabi University, India  
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India  
Dr. K.L. Shunmuganathan, R.M.K Engg College, Kavaraipeitai, Chennai  
Dr. Prasant Kumar Pattnaik, KIST, India.  
Dr. Ch. Aswani Kumar, VIT University, India  
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA  
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan  
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia  
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA  
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia  
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India  
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India  
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia  
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan  
Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA  
Mr. R. Jagadeesh Kannan, RMK Engineering College, India  
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India  
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh

Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India  
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia  
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India  
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India  
Dr. Ajay Goel, HIET , Kaithal, India  
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India  
Mr. Suhas J Manangi, Microsoft India  
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded , India  
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India  
Dr. Amjad Rehman, University Technology Malaysia, Malaysia  
Mr. Rachit Garg, L K College, Jalandhar, Punjab  
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India  
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan  
Dr. Thorat S.B., Institute of Technology and Management, India  
Mr. Ajay Prasad, Sir Padampat Singhanian University, Udaipur, India  
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India  
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh  
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia  
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India  
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA  
Mr. Anand Kumar, AMC Engineering College, Bangalore  
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India  
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India  
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India  
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow ,UP India  
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India  
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India  
Prof. Niranjana Reddy. P, KITS , Warangal, India  
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India  
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India  
Dr. A.Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai  
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India  
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan  
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India  
Dr. Tossapon Boongoen , Aberystwyth University, UK  
Dr . Bilal Alatas, Firat University, Turkey  
Assist. Prof. Jyoti Praaksh Singh , Academy of Technology, India  
Dr. Ritu Soni, GNG College, India  
Dr . Mahendra Kumar , Sagar Institute of Research & Technology, Bhopal, India.



Dr. Binod Kumar, India  
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan  
Dr. T.C. Manjunath , ATRIA Institute of Tech, India  
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan  
Assist. Prof. Harmunish Taneja, M. M. University, India  
Dr. Chitra Dhawale , SICSAR, Model Colony, Pune, India  
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India  
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad  
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India  
Mr. G. Appasami, Dr. Pauls Engineering College, India  
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan  
Mr. Yaser Miaji, University Utara Malaysia, Malaysia  
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh  
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India  
Dr. S. Sasikumar, Roever Engineering College  
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India  
Mr. Nwaocha Vivian O, National Open University of Nigeria  
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India  
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India  
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore  
Mr. Pawan Jindal, Jaypee University of Engineering and Technology, India  
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia  
Dr. Dhuha Basheer abdullah, Mosul university, Iraq  
Mr. S. Audithan, Annamalai University, India  
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India  
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India  
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam  
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India  
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad  
Mr. Deepak Gour, Sir Padampat Singhanian University, India  
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India  
Mr. Ali Balador, Islamic Azad University, Iran  
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India  
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India  
Dr. Debojyoti Mitra, Sir padampat Singhanian University, India  
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia  
Mr. Zhao Zhang, City University of Hong Kong, China  
Prof. S.P. Setty, A.U. College of Engineering, India  
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India  
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India

Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India  
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India  
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India  
Dr. Hanan Elazhary, Electronics Research Institute, Egypt  
Dr. Hosam I. Faiq, USM, Malaysia  
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India  
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India  
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India  
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan  
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India  
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia  
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India  
Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India  
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India  
Prof Anupam Choudhary, Bhilai School Of Engg., Bhilai (C.G.), India  
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya  
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.  
Dr. Kasarapu Ramani, JNT University, Anantapur, India  
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India  
Dr. C G Ravichandran, R V S College of Engineering and Technology, India  
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia  
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia  
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India  
Dr. Nikolai Stoianov, Defense Institute, Bulgaria  
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode  
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India  
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh  
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India  
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria  
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys  
research Group, Venezuela  
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India  
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia  
Mr. B. Muthu Kumar, Kathir College Of Engineering, Coimbatore  
Dr. Nighat Mir, Effat University, Saudi Arabia  
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India  
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore

Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore  
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US  
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India  
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India  
Mr. P. Sivakumar, Anna university, Chennai, India  
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia  
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India  
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia  
Mr. Nikhil Patrick Lobo, CADES, India  
Mr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India  
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India  
Assist. Prof. Vishal Bharti, DCE, Gurgaon  
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India  
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India  
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India  
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India  
Mr. Hamed Taherdoost, Tehran, Iran  
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran  
Mr. Shantanu Pal, University of Calcutta, India  
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom  
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria  
Mr. P. Mahalingam, Caledonian College of Engineering, Oman  
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt  
Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India  
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India

**CALL FOR PAPERS**  
**International Journal of Computer Science and Information Security**  
**IJCSIS 2011**  
**ISSN: 1947-5500**  
<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

***Track A: Security***

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity  
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

### ***Track B: Computer Science***

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail [ijcsiseditor@gmail.com](mailto:ijcsiseditor@gmail.com). Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .





**© IJCSIS PUBLICATION 2010**  
**ISSN 1947 5500**